

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2004-362559

(43)Date of publication of application : 24.12.2004

(51)Int.Cl.

G06F 13/00

(21)Application number : 2004-148159

(71)Applicant : MICROSOFT CORP

(22)Date of filing : 18.05.2004

(72)Inventor : GOODMAN JOSHUA T
ROUNTHWAITE ROBERT L
GWOZDZ DANIEL
MEHR JOHN D
HOWELL NATHAN D
RUPERSBURG MICAH C
STARBUCK BRYAN T

(30)Priority

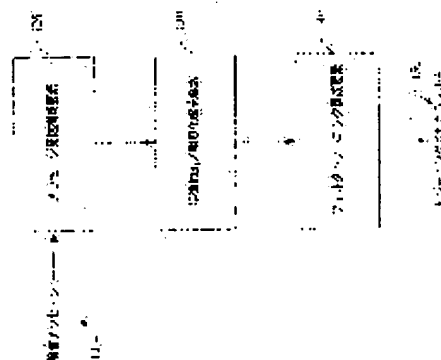
Priority number : 2003 454168 Priority date : 04.06.2003 Priority country : US

(54) FEATURES AND LIST OF ORIGATION AND DESTINATION FOR SPAM PREVENTION

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a system and method that facilitates extracting data from messages for spam filtering.

SOLUTION: The extracted data can be treated in the form of features, and the features can be used with machine learning system so that improved filters can be built. Origination information and such data as to be related to other information embedded in message body which enables a receipt of messages to get in touch and/or response to the sender of the messages can be extracted as features. The features or a subset can be deobfuscated and/or normalized, before they are used as features of the machine learning system. The features can be used to populate several feature lists which



facilitates prevention and detection of spams. Exemplary features includes an e-mail address, an IP addresses, a URL, embedded images indicating a URL and/or parts of them.

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2004-362559

(P2004-362559A)

(43) 公開日 平成16年12月24日(2004.12.24)

(51) Int. Cl.⁷
G06F 13/00

F I
G06F 13/00 610Q

テーマコード (参考)

審査請求 未請求 請求項の数 54 O L 外国語出願 (全 38 頁)

(21) 出願番号 特願2004-148159 (P2004-148159)
(22) 出願日 平成16年5月18日 (2004.5.18)
(31) 優先権主張番号 10/454,168
(32) 優先日 平成15年6月4日 (2003.6.4)
(33) 優先権主張国 米国 (US)

(71) 出願人 500046438
マイクロソフト コーポレーション
アメリカ合衆国 ワシントン州 9805
2-6399 レッドモンド ワン マイ
クロソフト ウェイ
(74) 代理人 100077481
弁理士 谷 義一
(74) 代理人 100088915
弁理士 阿部 和夫
(72) 発明者 ジョシュア ティー. グッドマン
アメリカ合衆国 98052 ワシントン
州 レッドモンド ノースイースト 38
ストリート 17424

最終頁に続く

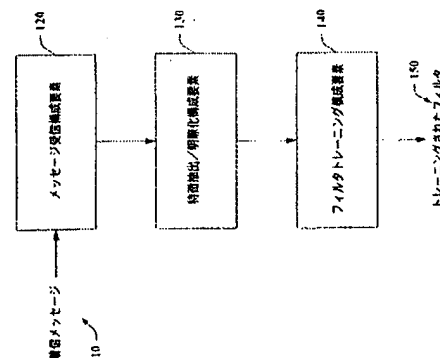
(54) 【発明の名称】 スпам防止のための送信元および宛先の特徴およびリスト

(57) 【要約】

【課題】 スпамフィルタリングのためにメッセージからデータを抽出することを容易にするシステムおよび方法を提供すること。

【解決手段】 抽出されたデータは特徴の形で扱うことができ、特徴は、機械学習システムと共に使用し、改良されたフィルタを構築することができる。送信元情報、ならびにメッセージの受信者がメッセージの送信者に接触および/または応答することを可能にするメッセージの本文内に埋め込まれた他の情報に関連付けられたデータの特徴として抽出することができる。特徴またはそのサブセットは、機械学習システムの特徴として使用する前に正規化および/または明瞭化することができる。特徴は、スパム検出および防止を容易にする複数の特徴リストに配置するために使用することができる。例示的な特徴には、電子メールアドレス、IPアドレス、URL、URLを指す埋込み画像、および/またはその一部分が含まれる。

【選択図】 図1



【特許請求の範囲】**【請求項1】**

スパム処理に関連するデータの抽出を容易にするシステムであって、

アイテムを受信し、メッセージの送信元もしくは該送信元の一部、および／または意図される受信者が該メッセージに関連して接触、応答、または受信することを可能にする情報に関連付けられた特徴のセットを抽出する構成要素と、

フィルタの構築に関連して、前記抽出された特徴のサブセットを使用する構成要素とを備えたことを特徴とするシステム。

【請求項2】

前記特徴のサブセットを明瞭化する正規化構成要素をさらに備えたことを特徴とする請求項1に記載のシステム。

【請求項3】

前記フィルタは、スパムフィルタであることを特徴とする請求項1に記載のシステム。

【請求項4】

前記フィルタは、ベアレンタル制御フィルタであることを特徴とする請求項1に記載のシステム。

【請求項5】

前記明瞭化された特徴を使用して、スパムおよび非スパムのうち少なくとも1つを学習する機械学習システム構成要素をさらに備えたことを特徴とする請求項1に記載のシステム。

【請求項6】

前記特徴のサブセットは、少なくとも1つのIPアドレスを含み、前記少なくとも1つのIPアドレスは、メッセージ内に位置する返信先アドレス、カーボンコピーアドレス、mailto:アドレス、発信元アドレス、およびURLのいずれか1つの少なくとも一部分であることを特徴とする請求項1に記載のシステム。

【請求項7】

前記IPアドレスは、ブロックIDを含み、前記ブロックIDは、少なくとも1つの特徴として抽出することができることを特徴とする請求項6に記載のシステム。

【請求項8】

前記ブロックIDは、少なくとも部分的には、ブロックディレクトリを調べることによって決定されることを特徴とする請求項7に記載のシステム。

【請求項9】

前記ブロックディレクトリは、arin.netであることを特徴とする請求項8に記載のシステム。

【請求項10】

前記ブロックIDは、少なくとも部分的には、推測することによって決定され、該決定により、前記IPアドレスの少なくとも最初の1ビット、少なくとも最初の2ビット、少なくとも最初の3ビット、および少なくとも最大で最初の31ビットまでのいずれか1つを特徴として抽出することを特徴とする請求項7に記載のシステム。

【請求項11】

前記特徴のサブセットは、IPアドレスの最初の1ビットから最初の31ビットまでの各々を含むことを特徴とする請求項1に記載のシステム。

【請求項12】

前記特徴のサブセットは、URLを含むことを特徴とする請求項1に記載のシステム。

【請求項13】

前記URLアドレスは、前記メッセージの本文内に配置されたもの、前記メッセージ内でテキストとして埋め込まれているもの、前記メッセージ内で画像として埋め込まれているもののうち少なくとも1つであることを特徴とする請求項12に記載のシステム。

【請求項14】

前記抽出された特徴の少なくともサブセットを使用して少なくとも1つの特徴リストに

配置する構成要素をさらに備えたことを特徴とする請求項1に記載のシステム。

【請求項15】

前記少なくとも1つの特徴リストは、問題のないユーザのリスト、スパム送信者のリスト、本物の送信者を示す肯定的な特徴のリスト、スパムを示す特徴のリストのいずれか1つであることを特徴とする請求項14に記載のシステム。

【請求項16】

前記特徴のサブセットは、少なくとも1つのURLを含むことを特徴とする請求項1に記載のシステム。

【請求項17】

前記URLは、前記メッセージの本文内にテキストとして埋め込まれていることを特徴とする請求項16に記載のシステム。

【請求項18】

前記URLは、前記メッセージの本文内のリンクの少なくとも一部分であることを特徴とする請求項16に記載のシステム。

【請求項19】

前記URLは、メッセージ内の画像として埋め込まれたリンクの少なくとも一部分であることを特徴とする請求項16に記載のシステム。

【請求項20】

前記特徴のサブセットは、電子メールアドレスから抽出されたホスト名およびドメイン名のうち少なくとも1つを含むことを特徴とする請求項1に記載のシステム。

【請求項21】

前記特徴のサブセットは、電子メールアドレスおよびURLのいずれか1つから抽出されたFQDNの少なくとも一部分を含むことを特徴とする請求項1に記載のシステム。

【請求項22】

前記特徴のサブセットは、電子メールアドレスおよびURLのいずれか1つから抽出されたドメイン名の少なくとも一部分を含むことを特徴とする請求項1に記載のシステム。

【請求項23】

前記抽出された前記特徴のサブセットの少なくとも一部分は、機械学習システムと共に使用される前に正規化されることを特徴とする請求項1に記載のシステム。

【請求項24】

前記抽出された前記特徴のサブセットの少なくとも一部分は、少なくとも1つの特徴リストに配置するために使用される前に正規化されることを特徴とする請求項1に記載のシステム。

【請求項25】

URL、電子メールアドレス、IPアドレスのうち少なくとも1つの少なくとも一部分を、成人、成人向け内容、ふさわしくない、一部の年齢にとってふさわしくない、あらゆる年齢にとってふさわしい、不適切、および適切なうちのいずれか1つとして分類する分類構成要素をさらに備えたことを特徴とする請求項1に記載のシステム。

【請求項26】

前記分類構成要素は、ペアレンタル制御システムであることを特徴とする請求項25に記載のシステム。

【請求項27】

前記分類構成要素は、少なくとも1つの特徴の型を、前記URL、ウェブサイトアドレス、前記IPアドレスのうち少なくとも1つの前記分類された一部分に割り当てることを特徴とする請求項25に記載のシステム。

【請求項28】

前記特徴のセットは、少なくとも1つの非無料電話番号を含み、前記電話番号は、前記メッセージに関連付けられた送信者または連絡先の地理的位置をマッピングすることを容易にするエリアコードを含むことを特徴とする請求項1に記載のシステム。

【請求項29】

請求項1のコンピュータ実行可能構成要素を格納することを特徴とするコンピュータ読取可能な媒体。

【請求項30】

請求項1に記載の前記システムを使用することを特徴とするコンピュータ。

【請求項31】

スパム処理に関連するデータの抽出を容易にする方法であって、

メッセージを受信するステップと、

前記メッセージの送信元もしくはその一部、および／または意図される受信者が前記メッセージに関連して接触、応答、または受信することを可能にする情報に関連付けられた特徴のセットを抽出するステップと、

フィルタを構築することに関連して、前記抽出された特徴のサブセットを使用するステップと

を備えたことを特徴とする方法。

【請求項32】

前記特徴のセットは、IPアドレスの少なくとも一部分を含むことを特徴とする請求項31に記載の方法。

【請求項33】

前記IPアドレスの少なくとも一部分を抽出するステップは、

ブロックIDが追加の特徴として抽出されるように、ブロックIDディレクトリを調べ、前記IPアドレスに対応する少なくとも1つのブロックIDを決定する動作、および

前記IPアドレスから少なくとも最初の1ビットから最初の31ビットまでの各々を抽出する動作のうち少なくとも1つを実行するステップを含むことを特徴とする請求項32に記載の方法。

【請求項34】

少なくとも1つの抽出されたIPアドレスは、少なくとも1つのサーバに対応することを特徴とする請求項32に記載の方法。

【請求項35】

前記少なくとも1つのサーバを追加の特徴として抽出するステップをさらに備えたことを特徴とする請求項34に記載の方法。

【請求項36】

前記メッセージから抽出された特徴の少なくともサブセットを明瞭化するステップをさらに備えたことを特徴とする請求項31に記載の方法。

【請求項37】

前記メッセージから抽出された少なくとも1つの特徴の少なくとも一部分を明瞭化するステップをさらに備えたことを特徴とする請求項31に記載の方法。

【請求項38】

前記メッセージから抽出された発信元IPアドレスを明瞭化するステップは、添付先(appended-to) IPアドレスの同一性を検証するため、複数の添付先IPアドレスを遡って追跡するステップを含むことを特徴とする請求項37に記載の方法。

【請求項39】

1度に少なくとも1つの接尾語を除去し、該除去により、それぞれの追加の特徴を生じる動作と、

1度に少なくとも1つの接頭語を除去し、該除去により、それぞれの追加の特徴を生じる動作とのうち、少なくとも1つを実行するステップを含む、ウェブサイトアドレスから追加の特徴を抽出するステップをさらに備えたことを特徴とする請求項37に記載の方法。

【請求項40】

前記特徴のセットは、返信先アドレス、カーボンコピーアドレス、mailto:アドレス、URL、リンク、および発信元アドレスのいずれか1つの少なくとも一部分を含むことを特徴とする請求項37に記載の方法。

【請求項41】

前記抽出された特徴の少なくともサブセットは、前記メッセージの本文内でテキストおよび画像のうち1つとして埋め込まれていることを特徴とする請求項31に記載の方法。

【請求項42】

前記特徴のセットは、ホスト名およびドメイン名を含むことを特徴とする請求項31に記載の方法。

【請求項43】

前記メッセージに関連付けられたふさわしい内容、およびふさわしくない内容のいずれか1つを示すため、1つまたは複数の抽出された特徴および／またはその一部分を分類するステップと、該分類を追加の特徴として使用するステップとをさらに備えたことを特徴とする請求項31に記載の方法。

【請求項44】

少なくとも部分的にはそれぞれの抽出された特徴に基づいて、メッセージ内容をユーザに通知するため、それぞれの抽出された特徴に特徴の型を割り当てるステップと、前記特徴の型を追加の特徴として使用するステップとをさらに備えたことを特徴とする請求項31に記載の方法。

【請求項45】

特徴の型および特徴の少なくとも1つが、希少性および共通性のいずれか1つであることを決定するステップと、特徴の希少性および共通性を追加の特徴として使用するステップとをさらに備えたことを特徴とする請求項44に記載の方法。

【請求項46】

前記特徴のサブセットは、機械学習システムを介してフィルタを構築することに関連して使用されることを特徴とする請求項31に記載の方法。

【請求項47】

前記フィルタは、スパムフィルタであることを特徴とする請求項31に記載の方法。

【請求項48】

前記フィルタは、ペアレンタル制御フィルタであることを特徴とする請求項31に記載の方法。

【請求項49】

前記メッセージから抽出された特徴の少なくともサブセットを使用して1つまたは複数の特徴リストに配置するステップをさらに含むことを特徴とする請求項31に記載の方法。

【請求項50】

特徴リストは、非スパム送信者を示す肯定的な特徴リスト、およびスパム送信者を示す悪質な特徴リストのうち少なくとも1つを含むことを特徴とする請求項49に記載の方法。

【請求項51】

前記抽出された特徴は、少なくとも部分的には機械学習システムの特徴として使用される前に、明瞭化されることを特徴とする請求項31に記載の方法。

【請求項52】

前記抽出された特徴は、少なくとも部分的には特徴リストに配置する特徴として使用される前に、明瞭化されることを特徴とする請求項31に記載の方法。

【請求項53】

メッセージからデータを抽出することを容易にして、複数のコンピュータプロセス間で送信されるように適合されたデータパケットであって、

メッセージを受信すること、前記メッセージの送信元もしくは該送信元の一部、および／または意図される受信者が前記メッセージに関連して接触、応答、または受信することを可能にする情報に関連付けられた特徴のセットを抽出することと、フィルタの構築に関連して前記抽出された特徴のサブセットを使用することとに関連付けられた情報を

備えたことを特徴とするデータパケット。

【請求項54】

スパム処理に関連するデータを抽出することを容易にするシステムであって、
メッセージを受信するための手段と、
前記メッセージの送信元もしくは該送信元の一部、および／または意図される受信者が
前記メッセージに関連して接触、応答、または受信することを可能にする情報に関連付け
られた特徴のセットを抽出するための手段と、
フィルタの構築に関連して、前記抽出された特徴のサブセットを使用する手段と
を備えたことを特徴とするシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、本物のメール（すなわち、問題のないメール）および望ましくないメールの
双方を識別するシステムおよび方法に関し、より詳細には、スパム防止を容易にするため
に、電子メッセージを処理してデータを抽出するためのシステムおよび方法に関する。

【背景技術】

【0002】

インターネットなど地球全体の通信ネットワークの登場により、膨大な数の潜在する顧
客に情報を届ける商業的な機会が提供された。電子メッセージング、特に電子メール（e
メール）は、（「スパム」とも呼ばれる）望まれてもいない広告および勧誘をネットワ
ークユーザにまき散らすための手段として、ますます広まりつつある。

【0003】

コンサルティングおよび市場調査会社であるRadicati Group, Inc.
は、2002年8月現在、毎日20億通のジャンク電子メールメッセージが送られており
、この数は2年毎に3倍になると予想されると見積もっている。個人および団体（たと
えば、企業、政府機関）はますます、ジャンクメッセージによって迷惑を受け、しばしば不
快な思いをさせられている。したがって、スパムは現在、またはまもなく、信頼できるコ
ンピューティングを揺るがす大きな脅威になるであろう。

【0004】

スパムを防ぐために使用される主な技法は、フィルタリングシステム／方法を使用する
ことである。実証済みのフィルタリング技法の1つは、機械学習手法に基づくものであり
、機械学習フィルタは、そのメッセージがスパムである確率を着信メッセージに割り当て
る。この手法では、一般に2種類の事例メッセージ（たとえば、スパムメッセージと非ス
パムメッセージ）から特徴（feature）が抽出され、この2つの種類間を確率的に
弁別するために学習フィルタが適用される。多数のメッセージの特徴は内容（たとえば、
メッセージの主題および／または本文内の単語または句）に関係するため、そのような型
のフィルタは、一般に「コンテンツベースのフィルタ」と呼ばれる。

【発明の開示】

【発明が解決しようとする課題】

【0005】

そのようなスパムフィルタリング技法の猛反撃を受け、多数のスパム送信者は、スパム
フィルタを回避および／または迂回するためにその正体を隠す方法を考えるようになって
いる。したがって、従来のコンテンツベースのフィルタや適応型フィルタは、偽装された
スパムメッセージを認識し遮断する上で効果を失うおそれがある。

【課題を解決するための手段】

【0006】

本発明のいくつかの態様についての基本的な理解を得るため、以下に本発明を簡単にま
とめる。この概要は、本発明の広範な全体像を示すものではない。本発明の主な／決定的
な要素を特定すること、あるいは本発明の範囲を説明することは意図されていない。後か
ら提供されるより詳しい説明の序文として、本発明のいくつかの概念を簡単な形態で示す
ことを目的とするにすぎない。

【0007】

スパム送信者は、そのメッセージ内のほとんどすべての情報を偽装することができる。たとえば、機械学習システムにとって特徴として使用される単語そのものが存在しないように、画像を埋め込むことができる。画像を所定の方法で歪ませて、OCRソフトウェアを使用することが困難になるか、または少なくとも時間がかかるようにすることもできる。しかし、どれだけ多数の特徴を除去しても、依然として有用な情報はある。第1に、スパム送信者は、どこからかメッセージを送らなければならない。どのIPアドレスからメッセージが届けられたか検出することができるのである。第2に、スパム送信者は大抵何かを販売しようとしており、したがって接触する方法が含まれているはずである。これは無料電話番号である可能性もあるが、スパム送信者は、苦情からコストが高つくため、これを使用するのを避けるかもしれない。無料でない電話番号である可能性もあるが、逆に応答率がより低いいため、そうしないかもしれない。またはこれに替えて、URLである可能性がある(たとえば、<http://www.spamcorp.com/buyenlarger.htm>)。このURLが、フィルタおよび/またはソフトウェアによって検出されるのをより困難にするために、画像内に埋め込まれている可能性がある。しかし、スパム送信者は、ユーザがURLをブラウザにタイプすることを必要とし、それにより応答率が低下する可能性があるため、そうしないであろう。

【0008】

スパム送信者が接触を受ける最も可能性の高い方法は、埋込みリンク、または何らかの埋込み電子メールアドレスによるものである。たとえば、「もっと知るにはここをクリックしてください」であり、「ここをクリック」には、本発明の一態様に従って機械学習システムが検出および使用することができる特定のウェブページへのリンクが含まれている。同様に、そこに返信すべきアドレス(たとえば、典型的には「発信元アドレス(from address)」であるが、「返信先(reply-to)」アドレスのある場合がある)、または任意の埋込みmailto:リンク(リンク上でクリックすることによってメールメッセージを送ることを可能にするリンク)、若しくは任意の他の埋込み電子メールアドレスである。さらに、スパム送信者は、しばしばメッセージ内に画像を含める。大きな画像を何回もメールするのはコストがかかるため、スパム送信者は、しばしばその画像への特別なリンクだけ埋め込み、そのリンクによってその画像のダウンロードが行われてしまう。これらのリンクが指す位置もまた、特徴として使用することができる。

【0009】

メール発信元アドレス、メール返信先アドレス、埋込みmailto:アドレス、外部リンク、および外部画像のリンクから引き出された情報に関して、そのような情報の少なくとも一部分を機械学習システムの特徴として使用することができ、重みまたは確率が関連付けられ、あるいは情報をリストに追加することができる。たとえば、スパムだけ、または問題のないメールだけ、または90%を超える問題のないメールなどを送信するIPアドレスまたは発信元アドレスのリストを保つことができる。特定のリンクまたはアドレスがそのようなリスト上にあるという事実を、機械学習システムの特徴として、または任意の他のスパムフィルタリングシステムの一部として、あるいはその両方として使用することができる。

【0010】

本発明は、メッセージの特定の一部分を調べることにより、偽装されたスパムメッセージを識別することを容易にするシステムおよび方法を提供する。より具体的には、本発明は、本物のメッセージからスパムメッセージを区別するために、電子メール(eメール)などメッセージを処理して送信元および/または宛先データを抽出するものである。この処理は、IPアドレス情報、電子メールアドレス情報、および/またはユニフォームリソースロケータ(URL)情報を識別して構文解析するための、および抽出されたデータをスパム属性(例えば、問題のないユーザと悪質なユーザ、または問題のない送信者と悪質な送信者)に関連付けるための様々な技法を含む。悪質なユーザまたは悪質な送信者は、たとえばスパム送信者(すなわち、スパムを送る者)と見なされるはずである。

【0011】

抽出されたデータ、または少なくともその一部分を使用し、機械学習システム用の特徴セットを生成することができる。機械学習技法は、メッセージの内容を調べ、そのメッセージがスパムであるかどうか判定する。スパム送信者は、処理するのが困難な画像内にその情報のほとんどを入れるなどのことをすることによって、メッセージの内容のほとんどを不明瞭にすることができる。しかし、スパム送信者は、受信者がスパム送信者に容易に接触する何らかの方法を提供することが必要なため、メッセージの送信源を完全に偽装することができない。そのようなものとしては、リンク（たとえば、URL）および／または電子メールアドレス（たとえば、IPアドレス）の使用などがある。これらの型の情報または変形形態、あるいはその一部分を、スパム検出器の特徴として使用することができる。具体的には、たとえば機械学習システムにより、その情報を使用してスパム検出器および／またはスパムフィルタをトレーニングすることができる。

【0012】

本発明はまた、ペアレנטラル制御システムと共に用いることができる。ペアレントラル制御システムは、メッセージが不適切であることをユーザに通知することができ、および「ポルノ題材を含む」など、その理由を示すこともできる。本発明の一態様によれば、1つまたは複数の抽出され正規化された特徴（たとえば、URL）をペアレントラル制御システムまたはフィルタに通し、ペアレントラル制御システムの分類を得ることができる。この分類を機械学習システムの追加の特徴として使用し、スパムフィルタを構築および／または改善することを容易にすることができる。

【0013】

さらに、抽出された特徴を種類によって分類し、スパム度に従って加重し、および肯定的（たとえば、スパムでない可能性がより高い）特徴または否定的（たとえば、スパムである可能性がより高い）特徴と指定することができる。また、特徴群を使用し、たとえば、非スパム送信者リストおよびスパム送信者リストなどのリストを作成することができる。

【0014】

前述の、および関連する目的を達成するため、本明細書では、本発明のある種の例示的な実施形態が、以下の説明および添付の図面と共に説明される。しかし、これらの実施形態は、本発明の原理を使用することができる様々な方法のいくつかを示すものにすぎず、本発明は、そのような実施形態とその等価なすべてを含むものとする。本発明の他の利点および新規の特徴は、図面と共に考察し、以下の本発明の詳細な説明から明らかにすることができる。

【発明を実施するための最良の形態】

【0015】

次に、本発明について図面を参照しながら説明する。図面では、全体を通して同じ要素を参照するために同じ符号が使用される。以下の説明では、あくまで説明の目的で、本発明を十分理解するための多数の具体的な詳細が説明される。しかし、これら具体的な詳細なしに本発明を実施することができることは自明であることが理解される。場合によっては、本発明の説明を容易にするため、周知の構造およびデバイスがブロック図の形態で示されている。

【0016】

本願では、「構成要素」および「システム」という用語は、ハードウェア、ハードウェアとソフトウェアの組合せ、ソフトウェア、または実行中のソフトウェアのいずれかであるコンピュータ関連のエンティティを指すものとする。たとえば、構成要素は、それだけには限らないが、プロセッサ上で動作するプロセス、プロセッサ、オブジェクト、実行可能物、実行のスレッド、プログラム、および／またはコンピュータとすることができる。例示のため、サーバ上で動作するアプリケーションもサーバと共に構成要素とすることができる。プロセスおよび／または実行のスレッド内に1つまたは複数の構成要素が常駐することができる。構成要素は、1つのコンピュータに局所化する、および／または複数のコンピュータ間で分散することができる。

【0017】

本発明は、機械学習式スパムフィルタリング用のトレーニングデータを生成することに関連して、様々な推理スキームおよび／または技法を組み込むことができる。本明細書では、「推理」という用語は、概して、事象および／またはデータを介して取り込まれた観察結果の集合から、システムの状態、環境、および／またはユーザについて推論する工程、またはこれらの状態を推理する工程を指す。推理は、たとえば、特定の内容または動作を識別するために使用することができ、あるいは状態全体にわたって確率分布を生成することができる。推理は、確率的なもの、すなわち、データおよび事象の考察に基づいた当該の状態全体にわたる確率分布の計算とすることができる。推理はまた、事象および／またはデータの集合から、より高いレベルの事象を構成するために使用される技法を指す可能性がある。そのような推理により、事象群が時間的に近接しているか否か、およびその事象およびデータが1つまたは複数の事象源やデータ源からのものであるか否かにかかわらず、観察された事象および／または記憶された事象データの集合から新しい事象または動作が構築される。

【0018】

本明細書全体にわたって「メッセージ」という用語が広く使用されているが、そのような用語は、電子メールそれ自体に限定されず、任意の好適な通信アーキテクチャを介して配布することができるどの形態の電子メッセージングをも含むように適切になすことができることを理解されたい。たとえば、2名以上での会議を容易にする会議アプリケーション（たとえば、対話型チャットプログラムおよびインスタントメッセージングプログラム）もまた、本明細書に開示されているフィルタリングの利点を活用することができる。というのは、望ましくないテキストは、ユーザがメッセージを交換しているとき、通常のチャットメッセージ内に電子的にばらまかれ、および／またはリードオフメッセージ、クロージングメッセージ、もしくは上記のすべてとして挿入される可能性があるからである。この特定のアプリケーションでは、望ましくない内容（たとえば、コマーシャル、勧誘、または広告）を取り込み、スパムとしてタグ付けするために、特定のメッセージ内容（テキストおよび画像）を自動的にフィルタリングするようにフィルタをトレーニングすることができるのである。

【0019】

本発明では、「受信者」という用語は、着信メッセージまたはメール項目の受取人を指す。「ユーザ」という用語は、状況に応じて、受信者または送信者を指す可能性がある。たとえば、ユーザは、状況や用語の適用に応じて、スパムを送信する電子メールユーザを指す可能性があり、および／またはスパムを受信する電子メール受信者を指す可能性がある。

【0020】

インターネットプロトコル（IP）アドレスは、一般にインターネット上の装置を表す32ビットの数値である。これらの数値は、2台の装置の間で通信するとき使用される。これらの数値は、一般に「xxx.xxx.xxx.xxx」の形態で表され、各xxxは、0と255の間である。残念ながら、IPアドレスは覚えておくことが困難である。そのため、「ドメイン名」とおよび「ホスト名」という取決めが作成されている。「ドメイン名」は、インターネット上の装置のグループ（おそらくは、単一の装置）の名前であり、一般に「x.com」または「y.edu」または「courts.wa.gov」という形態のものである。

【0021】

完全修飾ドメイン名（FQDN）は、インターネット上の特定の装置、たとえば「b.x.com」または「c.y.edu」または「www.courts.wa.gov」であり、ドメイン名部分は、それぞれ「x.com」または「y.edu」または「courts.wa.gov」である。「b」「c」「www」部分はそれぞれ、FQDNのホスト名部分と呼ばれる。一般に、IPアドレスは、ドメイン名を使用することができるどの状況でも使用することができる（たとえば、DN/IPは、どちらの可能性もあるこ

とを示す)。また、一般に、IPアドレスは、FQDNを使用することができる状況でも使用することができる(たとえば、FQDN/IPは、どちらの可能性もあることを示す)。電子メールアドレスは、たとえば「a@x.com」または「a@1.2.3.4」など、ユーザ名とドメイン名またはIPアドレス(DN/IP)からなる。どちらの例でも、ユーザ名は「a」である。

【0022】

ユニフォームリソースロケータ(URL)は、一般に「service-name:FQDN/IP/url-path」という形態のものである。たとえば、「http://www.microsoft.com/windows/help.htm」はURLである。「http」という部分がサービス名である。「www.microsoft.com」という部分はFQDNであり、「windows/help.htm」はURLパスである。これはURLを幾分簡単にしたものであるが、本考察には十分である。

【0023】

次に図1を参照すると、本発明の一態様による特徴抽出およびトレーニングシステム100の全体的なブロック図が示されている。特徴抽出およびトレーニングシステム100は、着信メッセージ110を処理し、そのメッセージからデータまたは特徴を抽出するものである。そのような特徴は、メッセージおよび/またはその変形の形態において提供される送信元および/または宛先情報の少なくとも一部分から抽出することができる。具体的には、1つまたは複数の着信メッセージ110は、メッセージ受信構成要素120を介してシステム100によって受信することができる。メッセージ受信構成要素120は、たとえば、着信メッセージ110を受信するために、電子メールサーバまたはメッセージサーバ上に位置することができる。一部のメッセージ(たとえば、少なくとも1つ)は既存のフィルタ(たとえば、スパムフィルタ、ジャンクメールフィルタ、ペアレナタル制御フィルタ)に弱く、したがって、ゴミ箱またはジャンクメールフォルダに転送される可能性があるが、送信元および/または宛先データの少なくとも一部分は、機械学習システムと共に、あるいは特徴リストに配置するに際して使用するために、抽出し明瞭にすることができる。

【0024】

メッセージ受信構成要素120は、着信メッセージ、またはそのサブセットを特徴抽出構成要素130に渡すことができる。特徴抽出構成要素130は、フィルタトレーニング、最終的にはスパム検出を容易にするための特徴セットを生成するために、それぞれのメッセージ110からデータを抽出することができる。メッセージから抽出されたこのデータまたは特徴は、メッセージ内で見つけられた、および/または埋め込まれた送信元および/または宛先情報に関連する。データまたは特徴の例には、発信元(received-from) IPアドレス、返信先(reply-to)電子メールアドレス、cc:(たとえば、カーボンコピー)電子メールアドレス、各種URL(テキストをベースとするリンク、画像をベースとするリンク、およびテキスト形態のURLまたはその一部分)、非無料電話番号(たとえば、特にエリアコード)、無料電話番号、mailto:電子メールアドレスリンク、テキスト形態の電子メールアドレス、SMTP HELOコマンド内のFQDN、SMTP MAIL FROMアドレス/リターンパスアドレス、および/または上記のいずれかの少なくとも一部分が含まれる。

【0025】

特徴抽出構成要素130は、任意の適切な数のプロセスを実行し、後に機械学習システムで使用するために、メッセージ110から様々な特徴セットを抽出することができる。加えて、またはこれに代えて、特徴セットは、他のフィルタトレーニング技法用のリストに配置させるために使用することができる。

【0026】

たとえば、a.x.comなどFQDNは、一般にIPアドレスと呼ばれる番号に変換することができる。IPアドレスは、一般に、4つの番号ブロックを備えたドット付きの10進フォーマットとされる。各ブロックは、点または小数点によって分離され、各番号

ブロックは、0から255の範囲に及ぶことができ、番号の各変化は、異なるインターネット名に対応する。たとえば、a. x. comは123. 124. 125. 126に変換される可能性があり、一方、121. 124. 125. 126はq r s t u v. comを表す可能性がある。番号は、単語ほど容易に認識または記憶することができないため、IPアドレスは通常、それぞれのFQDNによって参照される。また、ドット付き10進フォーマットの同じIPアドレスを、以下で説明されることになる代替フォーマットで表すこともできる。

【0027】

本発明の一態様によれば、特徴抽出構成要素130は、メッセージ110内に含まれる発信元IPアドレスに集中して処理することができる。発信元IPアドレスは、少なくとも部分的には、発信元IP情報に基づくものである。一般に、インターネットを介して送信されたメールは、サーバからサーバに移送され、時に2つのサーバ（たとえば、送信側と受信側）を必要とするだけである。まれではあるが、クライアントがサーバに直接送信することができる。場合によっては、たとえば、ファイアウォールがあるために、メールまたはメッセージが、あるサーバから別のサーバに送信される。具体的には、一部のサーバは、ファイアウォールの内側に位置する可能性があり、したがって、そのファイアウォールの他方の側の指定されたサーバと通信することができるだけである。これは、送信側から受信側に到達するためにメッセージが取るホップの数の増加を引き起こす。そのIPアドレスを含む発信元の行により、メッセージがどこから来たか突き止めるため、メッセージのパスを追跡することが容易になる。

【0028】

メッセージ110がサーバからサーバへ移動するにつれ、途中の各サーバは、伝送されているサーバの主張しているFQDNの名前に加えて、メッセージが届けられたIPアドレスの識別をメッセージの発信元フィールド（すなわち、「Received:」フィールド）の先頭に付加する。このFQDNは、SMTPプロトコルのHELOコマンドを介して送信側サーバによって受信側サーバに伝送され、したがって、送信側サーバがそのシステムの外側にある場合には信頼することができない。たとえば、メッセージは、5つのIPアドレスとFQDNが付加された5つの発信元行を有し、したがって、行が付加された逆順（すなわち、最後が先頭）の状態で、メッセージが6つの異なるサーバを通った（すなわち、5回通過した）ことを示すことができる。しかし、各サーバは、より下方の（より早く付加された）行を修正することができる。これは、特にメッセージが複数のサーバ間を移動してきたとき、特に問題をはらむこととなる。各中間サーバは、より早く書き込まれた任意の（より下方の）発信元行を変えることが可能であるため、スパム送信者は、発信元IP情報またはスパムメッセージの送信者を偽装するため、メッセージの発信元の行に偽のIPアドレスを付加することができる。たとえば、スパムメッセージは、当初、trusted domain. com（信頼できるドメイン）から送信されたものであるかのように見え、したがって、メッセージの真の送信源について受信者に偽りを伝える可能性がある。

【0029】

スパムソフトウェアにとって、システムの内側のサーバに送信したシステム外側のIPアドレスを容易に識別することは重要である。このIPアドレスは、システム内側の受信サーバによって書き込まれるため、正しいIPアドレスとして信頼され得る。システム外側の他のIPアドレスはすべて、システム外側のサーバによって書き込まれており、したがって、修正されているおそれがあるため、信頼されない可能性がある。受信者システムへのパス内に送信サーバの多数のIPアドレスが含まれる可能性があるが、1つを信頼することができるにすぎないため、この信頼できるIPアドレスを「送信者の」IPアドレスと呼ぶものとする。

【0030】

スパムフィルタリングソフトウェアがこの送信者のIPアドレスを見つけるための1つの方法としては、システムでのメールサーバ構成を知ることが挙げられる。一般に、どの

状況でどの装置がどの他の装置に渡すか知っている場合、送信者のIPアドレスを決定することができる。しかし、特に電子メールクライアントにインストールされたスパムフィルタリングソフトウェアにとって、サーバ構成を記述することは都合が悪い場合がある。これに代わる手法には、MXレコードを使用してメッセージの真の送信源を決定するステップが含まれる。MXレコードは、各ドメイン名ごとに、そのドメインについて電子メールの受信者のFQDNをリストする。そのドメインのMXレコード内のエントリに対応するFQDNに対応するIPアドレスが見つかるまで、発信元リストを遡って追跡することができる。この装置が受信した発信元のIPアドレスは、送信者のIPアドレスである。

1. 2. 3. 101がx. comについて唯一のMXレコードであると想像してみる。次いで、1. 2. 3. 101から届けられた行を見つけることにより、次の行がx. comの着信メールサーバに対応すること、したがって、その行内のIPアドレスが、x. comに送信したIPアドレスに対応することを知ることができる。

【0031】

下記の表は、メッセージの真の送信源を決定する上記で説明した例示的な解析を示す。

【0032】

【表1】

行	解釈
Received: from a.x.com ([1.2.3.100]) by b.x.com Tue, 22 Apr 2003 13:11:48 -0700	x.comの内部
Received: from mailserver.x.com ([1.2.3.101]) by b.x.com Tue, 22 Apr 2003 12:11:48 -0700	1. 2. 3. 101は、x.comについてのMXレコードであり、それにより、次の行がx.com内部の最初であることがわかる。
Received: from outside.com ([4.5.6.7]) by mailserver.x.com Tue, 22 Apr 2003 11:11:48 -0700	これは、x.comがメッセージを受信した所である。すなわち、これは最後の信頼できる行である。4. 5. 6. 7を送信者のIPアドレスとして使用する。
Received: from trustedsender.com ([8.9.10.11]) by outside.com Tue, 22 Apr 2003 10:11:48 -0700	この行は、4. 5. 6. 7でサーバによって構築された偽物である可能性がある。

【0033】

現在、発信メールサーバをリストするための受け入れられている標準はなく、たとえば、システム内部のIPアドレスがシステム外部のIPアドレスと異なる場合、あるいは、システムが、MXレコード内にリストされているある装置からMXレコード内にリストされている別の装置に間接的に送信する場合、このヒューリスティックは失敗する可能性がある。さらに、MXレコード内のある装置がMXレコード内の別の装置に送信した場合に発生する可能性があるように、上記のように見つけれられた送信者のIPがそのシステムの内部にあると判明した特別な場合、プロセスは上記のように継続される。さらに、ある種のIPアドレスを内部として検出することができる（というのは、内部IPアドレスのためだけに使用されている形態である、10. x. y. zまたは172. 16. y. zから172. 31. y. zまたは192. 168. 0. zから192. 168. 255. zの形態のものであるからである）。すなわち、システム内部のどのアドレスも信頼することができる。最後に、発信元の行が「Received from a.x.com [1. 2. 3. 100]」という形態のものであり、a.x.comのIPアドレスルックアップが1. 2. 3. 100を生じる、あるいは、1. 2. 3. 100の逆IPアドレスルックアップがa.x.comを生じる場合、また、x.comがそのシステムである場合には、次の行もまた信頼することができる。

【0034】

これらの観察結果を使用して、送信者のIPアドレスを見つげられることが多い。例示的な疑似コードは次の通りである。

【0035】

【表2】

```
bool fFoundHostInMX;  
if (external IP address of MX records matches internal IP  
address of MX records)  
{
```

【0036】

【表3】

```
fFoundHostInMX = FALSE; # it's worth looking for
} else {
    fFoundHostInMX = TRUE; # it's not worth looking for,
    pretend we already found it
}

for each received from line of the form Received from a.b.c
[i.j.k.l] {
    if i.j.k.l in MX records of receiver domain
    {
        fFoundHostInMX = TRUE;
        continue;
    }
    if not fFoundHostInMX
    {
        # Has not yet gone through an MX record, must be
internal
        continue;
    }
    if i.j.k.l is of form
        10.x.y.z or
        172.16.y.z to 172.31.y.z or
        192.168.0.z to 192.168.255.z
    {
        # Must be internal
        continue;
    }
    if DNS lookup of a.b.c yields i.j.k.l and b.c is
receiver domain
    {
        # Must be internal
        continue;
    }
}
```

【0037】

【表4】

```

}

Output sender's alleged FQDN a.b.c and sender's actual
IP address i.j.k.k

}

If we reach here, then Error: unable to identify sender's
alleged FQDN and sender's actual IP address

```

【0038】

他の送信元および宛先機能の場合と同様に、送信者のIPアドレスを用いて多くの処理を行うことができる。第1に、ブラックリストと呼ばれることもある一様に悪質な送信者のリストに追加することができる。ブラックリストは、信頼できないメッセージをフィルタし、遮断し、または、それらをさらに調査することができる適切なフォルダまたは位置に向けて送り直すために後に使用することができる。

【0039】

また、他の型のリストを生成し、クライアントベースのアーキテクチャとサーバベースのアーキテクチャのどちらでもフィルタとして実行することができる。クライアントアーキテクチャでは、ユーザは（たとえば、メーリングリスト、個人など）誰からのメールを受信するかをクライアント電子メールソフトウェアに通知することができる。信頼される電子メールアドレスに対応するレコードのリストを、ユーザが手動で、または自動的に生成することができる。したがって、電子メールアドレス「b@zyx.com」を有する送信者がユーザに電子メールメッセージを送信すると考える。送信者の電子メールアドレスb@zyx.comは、ユーザ名「b」とFQDN/IP「zyx.com」を含む。クライアントが送信者（b@zyx.com）から着信メッセージ110を受信すると、そのユーザの電子メールアドレスについて信頼される送信者リストを探索し、「b@zyx.com」が有効かつ信頼されるアドレスであることをユーザが示しているかどうか判定することができる。サーバアーキテクチャの場合、リストは、サーバ上に直接配置することができる。したがって、メッセージがメッセージサーバに到着したとき、それぞれの特徴（たとえば、送信者のIPアドレス、MAIL FROMまたはHELOフィールド内のドメイン名、ならびに他の送信元および/または宛先情報）をメッセージサーバ上に配置されるリストと比較することができる。有効な送信者からのものと決定されたメッセージは、クライアントベースの送達プロトコルまたはサーバベースの送達プロトコルに従って、意図された受信者に送達することができる。しかし、疑わしいまたは悪質な特徴のリスト内の送信元または宛先の特徴を含むと決定されたメッセージは、廃棄するため、または他の方法で特別に処理するため、スパムまたはジャンクメールフォルダに移動することができる。

【0040】

信頼される送信元の特徴または悪質な送信元の特徴のリストに配置することに代わる方法として、送信者の送信元の特徴（たとえば、IPアドレス、主張されている発信元アドレス）を1つまたは複数の特徴として抽出し、フィルタ構築および/またはトレーニングのために機械学習技法と共に後に使用することができる。

【0041】

IPアドレスは、メッセージヘッダの任意の一部内の電子メールアドレス（たとえば、送信者のアドレスまたは返信先アドレス内のFQDN上のIPルックアップ）から、または、メッセージの本文内に埋め込まれたURLリンクのドメイン名部分のIPアドレスルックアップから、またはIPアドレスがURLのFQDN/IP部分として見られる場合には直接そこから導出することができる。さらに、後に説明するように、IPアドレスはいくつかの属性を有し、その各々を、機械学習システムの特徴として、またはユーザによ

ってボピュレートされるリストの要素として使用することができる。したがって、第2の手法では、特徴抽出構成要素130は、IPアドレスの多数の下位区分を利用し、追加の特徴を生成することができる。

【0042】

上記の特徴の任意の組合せは、各着信メッセージ110から抽出することができる。典型的にはメッセージすべてを使用することができるが、メッセージは、ランダムに、自動的に、および／または手動で選択し、特徴抽出に加わることができる。抽出された特徴セットは、機械学習システム、あるいはスパムフィルタなどフィルタ150を構築および／またはトレーニングする任意の他のシステムなど、フィルタトレーニング構成要素140に後に加えられる。

【0043】

図2を参照すると、本発明の一態様による着信メッセージ210の1つまたは複数の特徴を明瞭化(deobfuscate)または正規化することを容易にする特徴抽出システム200が示されている。最終的には、少なくとも部分的には正規化された特徴の1つまたは複数に基づいて、フィルタを構築することができる。システム200は、たとえば、図のように直接、またはメッセージ受信側(図1)によって間接的に着信メッセージ210を受信する特徴抽出構成要素220を含む。特徴抽出のために選択された、または特徴抽出に加わる着信メッセージは、ユーザ嗜好に従ってシステム200の対象とすることができる。またはこれに代えて、実質的にすべての着信メッセージを特徴抽出のために使用可能にし、また特徴抽出に加えることができる。

【0044】

特徴抽出は、メッセージ210からの送信元および／または宛先情報に関連付けられた1つまたは複数の特徴230(特徴₁232、特徴₂234、特徴_M236とも呼ばれ、ただし、Mは1以上の整数である)を引き出すことが必要である。送信元情報は、メッセージの送信者を示す要素、ならびにサーバドメイン名に関連し、およびメッセージが来た所を指定する識別情報に関連する可能性がある。宛先情報は、受信者がそのメッセージに対する応答を誰に、またはどこに送信することができるかを示すメッセージの要素に関連する可能性がある。送信元および宛先情報は、メッセージ受信者に見える、または見えない状態で(たとえば、テキストとして、または画像内に埋め込まれて)、メッセージのヘッダ内に、ならびにメッセージの本文内に見いだすことができる。

【0045】

スパム送信者は、従来のスパムフィルタによって検出されるのを回避するために正体を偽装し、および／または不明瞭化しようとする傾向がよくあるので、システム200は、1つまたは複数の抽出された特徴230、またはその少なくとも一部分を明瞭化するようにすることを促進する特徴正規化構成要素240を備える。特徴正規化構成要素240は、抽出された特徴230(たとえば、FQDNであるが、これは、ブロックおよびMXレコードのディレクトリを調べ、および／またはその現行フォーマットに従って変換される)を解析し、次いでそれらと、たとえば既存のスパム送信者リスト、非スパム送信者リスト、および／またはベアレンタル制御リストのデータベースとを比較することなどにより、抽出された特徴230を処理および／または分析することができる。図4において上記で説明したいくつかの場合には、抽出された特徴がURLであるときなど、接頭語および／または接尾語を除去して特徴を正規化すること、およびそのURLがスパム送信者のウェブサイトを指しているか、あるいは本物の送信源を指しているかを識別することを容易にすることもできる。

【0046】

特徴が正規化された後に、機械学習システムなどトレーニングシステム260がそれらの少なくともサブセット250を使用し、フィルタ270を構築および／または更新することができる。フィルタは、たとえばスパムフィルタおよび／またはジャンクメールフィルタとして使用するためにトレーニングすることができる。さらに、非スパム源(たとえば、送信者の発信元電子メールアドレス、送信者のIPアドレス、埋込み電話番号、およ

び／またはURL)および／または非スパム送信者を示すものなどの肯定的な特徴によって、ならびにスパム送信者を識別したり、スパム送信者に関連付けられたりするものなどの否定的な特徴によってフィルタを構築および／またはトレーニングすることができる。

【0047】

またはこれに代えて、あるいはこれに加えて、特徴セットは、新しいスパム特徴リスト280に配置するために、または既存のスパム特徴リスト280に追加するために使用することができる。また、問題のないアドレスのリスト、悪質なアドレスのリスト、問題のないURLのリスト、悪質なURLのリスト、問題のない電話番号のリスト、悪質な電話番号のリストなど、特定の抽出された特徴に対応するように他のリストを生成することができる。問題のない特徴リストは、非スパム送信者、履歴からみて本物の送信者、および／または非スパム送信者である尤度がより高い送信者（例えば、スパム源でない公算が90%以下）を識別することができる。逆に、悪質な特徴リストは、スパム送信者、潜在的なスパム送信者、および／またはスパムである尤度が比較的高い送信者（たとえば、90%以下のスパム源）に対応付けすることができる。

【0048】

次に図3～6は、本発明のいくつかの態様に従ってスパム検出および防止を容易にするため、IPアドレス、FQDN、電子メールアドレス、およびURLからそれぞれ導出または抽出することができる例示的な特徴を示す。

【0049】

図3は、本発明の態様によるIPアドレス300の例示的な内容を示す。IPアドレス300は32ビット長であり、ドット付き10進フォーマット（たとえば、それぞれ3桁までの4ブロックであり、各ブロックはピリオドによって分離され、3桁の各ブロックは、0と255の間で可分の任意の数である）で表されたとすると、ブロック（たとえば、ネットブロック）内に割り振られる。ブロックは、クラスA、クラスB、クラスCなど、クラスに割り当てられる。各ブロックは、IPアドレスの複数のセットを含み、ブロック当たりのIPアドレス数は、クラスによって変わる。すなわち、クラス（すなわち、A、B、またはC）に応じて、ブロック当たり、より多くの、またはより少ないアドレスが割り当てられる可能性がある。ブロックサイズは、通常、2のべき乗であり、同じブロック内のIPアドレスのセットは、最初のk個の2進数字を共有し、最後の $32-k$ （ 32 引く k ）個の2進数字が異なることになる。したがって、各ブロックは、その共有されている最初のk個のビットに従って識別することができる（ブロックID302）。特定のIPアドレス300に関連付けられたブロックID302を決定するために、ユーザは、arin.netなどブロックのディレクトリを調べることができる。さらに、ブロックID302は、特徴として抽出および使用することができる。

【0050】

しかし、場合によっては、ブロック内のIPアドレスのグループが分割されて処分され、また任意の回数だけ使いまわされる可能性があるため、arin.netを参照してもブロックID302を容易に決定することができない。そのような場合には、ユーザまたは抽出システムは、各々のIPアドレスについて1回または複数回、ブロックID302を推測することができる。たとえば、ユーザは、機械学習システムによって後に使用するための別の特徴として、および／または特徴リスト（たとえば、問題のない特徴リスト、スパム特徴リストなど）上の要素として、少なくとも最初の1ビット304、少なくとも最初の2ビット306、少なくとも最初の3ビット308、少なくとも最初のMビット310（すなわち、Mは1以上の整数）、および／または少なくとも最初の31ビット312まで抽出することができる。

【0051】

実際には、たとえば、IPアドレスの最初の1ビットを特徴として抽出および使用し、IPアドレスがスパム送信者を指しているか、あるいは非スパム送信者を指しているか判定することができる。他のメッセージから抽出された他のIPアドレスからの最初の1ビットを比較し、少なくとも1つのブロックIDを決定することを容易にすることができる。

のである。次いで、少なくとも1つのブロックを識別することは、そのメッセージがスパム送信者からのものかどうか見分けるのに役立つ可能性がある。さらに、最初のMビットを共有するIPアドレス群をそれらの他の抽出された特徴について比較し、IPアドレスが本物の送信者からのものかどうか、および/またはメッセージがそれぞれスパムかどうか突き止めることができる。

【0052】

また、IPアドレスを階層(314)によって構成することができる。すなわち、1組のより高次のビットを特定の国に割り振ることができる。その国では、ISP(インターネットサービスプロバイダ)を一定のサブセットに割り振り、次いで、そのISPでは、特定の会社を一定のサブセットに割り振ることができる。したがって、同じIPアドレスでも様々なレベルをもつことに意義がある。たとえば、IPアドレスが韓国に割り振られたブロックから来たという事実は、そのIPアドレスがスパム送信者に関連付けられているかどうか判定する上で有用である可能性がある。そのIPアドレスが、スパム送信者に対して厳しいポリシーを有するISPに割り振られたブロックの一部である場合、これもまた、そのIPアドレスがスパム送信者に割り当てられていないと決定する上で有用である可能性がある。したがって、IPアドレスの最初の1~31ビットの各々を、IPアドレス群の少なくともサブセットの階層構成314と組み合わせて使用することにより、ユーザは、IPアドレスが割り振られた方法を実際に知ることなしに(たとえば、ブロックIDを知ることなしに)、様々なレベルで自動的に情報を学習することができる。

【0053】

上記で説明した特徴に加えて、特徴の稀少性316(たとえば、特徴の発生があまり普通でない)は、たとえば、適切な計算を実行すること、および/または着信メッセージをサンプリングする際に特徴が現れる周期またはカウントを比較する統計データを使用することによって決定することができる。実際には、めったにないIPアドレス300が、電子メールを送達するために使用されるダイヤルアップ回線である可能性があるが、これはスパム送信者によってしばしば使用される戦法である。スパム送信者は、その識別および/または位置を頻繁に修正する傾向がある。したがって、その特徴は普通である、またはめったにないという事実は有用な情報となる。したがって、特徴の稀少性316は、機械学習システムの特徴として、および/または少なくとも1つのリスト(たとえば、稀少特徴リスト)の一部として使用することができる。

【0054】

図4は、たとえばb.x.comなどのFQDN400の例示的な特徴の内容を示す。FQDN400は、たとえば、HELOフィールドから抽出することができ(たとえば、送信者であると主張するFQDN)、一般に、ホスト名402およびドメイン名404を含む。ホスト名402は、その例によれば「b」である特定のコンピュータを指す。ドメイン名404は、インターネット上の少なくとも1つの装置または装置のグループの名前を指す。本例では、「x.com」は、ドメイン名404を表す。FQDN400の階層の内容は、内容406によって表される。具体的には、B.X.COM408(完全なFQDN400)は、部分的にX.COM410(部分的なFQDN)に分解することができ、次いで、これをCOM412(部分的なFQDN)に分解することができ、それにより、各部分的FQDNを特徴として使用することができる。

【0055】

発信元情報など、いくつかの特徴は、主にIPアドレスとして存在する。したがって、FQDN400を、(図3に示すように)IPアドレス300に変換し、追加の特徴として分析することは有用である可能性がある。というのは、新しいホスト名やドメイン名を作成するのは比較的容易であるが、新しいIPアドレスを得るのは比較的困難だからである。

【0056】

残念ながら、ドメインの所有者は、明らかに異なる装置すべてを同じ場所にマッピングさせることができる。たとえば、「a.x.com」という名前の装置の所有者は、「x

「.com」の同じ所有者である「b.x.com」の所有者と同じである可能性がある。したがって、スパム送信者は、そのメッセージが、ドメイン404「x.com」からのものではなく、FQDN400「b.x.com」からのものであると信じるように従来型フィルタを容易に欺き、それにより、実際には、そのメッセージがスパムであった、またはスパムである可能性がより高いことをドメイン404「x.com」が示していたかもしれない場合でも、メッセージがスパムフィルタを通過させてしまう可能性がある。したがって、メッセージの送信元および／または宛先情報を抽出するとき、アドレスを単にドメイン名404に分解することは有用である。またはこれに代えて、またはこれに加えて、完全FQDN400を特徴として抽出することができる。

【0057】

システムによっては、ベアレンタル制御システムなど追加の資源が使用可能な場合がある。これらの資源は、しばしば、ポルノや暴力など「型」または質的評価をホスト名、および／またはURLに割り当てることができる。抽出された特徴は、そのような資源を使用して、型によってさらに分類することができる。次いで、改良されたスパム関連フィルタを構築すること、および／またはトレーニングすることと併せて、特徴の型414を追加の特徴として使用することができる。またはこれに代えて、先に識別されている様々な特徴の型に対応して、リストを生成することができる。特徴の型414には、それだけに限らないが、性またはポルノ関連の特徴、人種および／または憎悪発言関連の特徴、肉体強化の特徴、収入または金銭的解決策の特徴、住宅購入の特徴などが含まれ、これらは、メッセージの一般的な目的を識別する。

【0058】

最後に、特徴の稀少性316、または特徴の型の希少性（上記図3参照）を、図3において上記で説明した別の特徴とすることができる。たとえば、FQDN400「b.x.com」のホスト名「B」402など、メッセージから抽出された特徴を特徴の型に共通する事例、例えばポルノ題材とすることができる。したがって、この特徴がメッセージから抽出され、次いでポルノ題材特徴リスト上で見つかったときには、そのメッセージは、スパムである可能性がより高い、またはあらゆる年齢にとってふさわしくない／不適切である、若しくは成人向け内容（たとえば、成人レーティング）を構成する、などと結論を下すことができる。したがって、各リストは、その特定の型の、より共通（common）の特徴を含むことができる。またはこれに代えて、一般に、対応するIPアドレスがスパムメッセージ内で共通して見出され、したがってスパムの共通の特徴として指定することができる。さらに、特徴の共通性（commonality）および／または希少性を、装置学習または他の規則をベースとするシステム用の別の特徴として使用することができる。

【0059】

図5は、FQDN400ならびにユーザ名502などいくつかの追加の特徴を含む、a@b.x.comという電子メールアドレス500の例示的な特徴の内容を示す。電子メールアドレス500は、メッセージのFromフィールド、cc（カーボンコピー）フィールド、reply-toフィールドから、ならびに、メッセージ本文内のmailto：リンクのいずれか（たとえば、mailto：リンクは、クリックされたとき特定のアドレスへのメールを生成する特別な種類のリンクである）から、さらに使用可能な場合、SMTPプロトコル内で使用されるMAIL FROMコマンドから抽出することができる。また、電子メールアドレス500は、メッセージの本文内にテキストとして埋め込むことができる。場合によっては、メッセージ内容は、そのメッセージに応答するとき「reply all（全員に返信）」機能を使用するように受信者を導くものである可能性がある。そのような場合には、ccフィールド内のアドレス、および／または「to」フィールド内に含まれるアドレスの少なくとも一部分（複数の受信者がリストされている場合）もまた、返信先であろう。したがって、これらのアドレスのそれぞれを1つまたは複数の特徴として抽出し、スパム送信者識別／防止を容易にすることができるであろう。

【0060】

「a@b. x. com」という電子メールアドレス500は、様々な要素または下位区分に分析することができ、これらの要素もまた、特徴として抽出および使用することができる。具体的には、電子メールアドレスは、ユーザ名502と、さらに追加の特徴に細分化することができるFQDN504（たとえば、図4のFQDN400参照）とを含む。使用、認識、および想起の容易さなど、いくつかの実際的な理由により、電子メールアドレスは、通常、IPアドレスではなくFQDNを使用して表記される。

【0061】

この例では、「a@b. x. com」は、ユーザ名502「a」を含む。したがって、「a」を1つの特徴として抽出することができる。同様に、FQDN504「b. x. com」を、少なくとも1つの他の特徴として電子メールアドレスから抽出することができる。電子メールアドレス500のFQDN504部分は、図4において上記により詳しく説明した特徴の型414を決定することを容易にするため、ペアレンタル制御フィルタに通することができる。したがって、電子メールアドレス500のFQDN部分に関する特徴の型を追加の特徴として使用することができる。

【0062】

電子メールアドレスに加えて、スパム送信者は、URLを介してアクセスを受けることがよくある。図6は、本発明の態様による例示的なURL600（たとえば、x. y. com/a/b/c）と、そこから抽出された複数の特徴とを示す。URL600は、メッセージ本文内のテキストとして、および／またはメッセージ本文内の画像として埋め込まれる可能性がある。たとえば、スパムメッセージは、ウェブサイトに対するポインタを含み、それにより、受信者をスパム送信者のウェブサイトまたは関連サイトに導く可能性がある。

【0063】

URLは、IPアドレスについてしたと同様な方法で明瞭化することができる。URL600を明瞭化する前に、まず、たとえば、http://、https://、ftp://、telnet://など任意の接頭語（たとえば、サービス名）を除去することができる。さらに、「@」記号（たとえば、16進表記で%40）がURLの中間に現れた場合、接頭語（たとえば、http://）と「@」記号の間にどんなものがあったとしても除去することができ、それからURL600を正規化することができる。スパム送信者による別の戦法または策略の形態として、接頭語と「@」記号の間にテキストを組み込むことにより、受信者が誘導されている真のページ位置に関してメッセージ受信者を混乱させることも考えられる。

【0064】

たとえば、http://www.amazon.com@121.122.123.124/info.htmは、メッセージ受信者にとって、このページがwww. amazon. comであるかのように見える。したがって、受信者がそのリンクを、およびより重要なことには、そのメッセージ送信者を信頼する傾向がより強くなる可能性がある。それに反して、真のページ位置は、実際にスパム関連ウェブページに対応する「121. 122. 123. 124」にある。しかし、場合によっては、自動ログインを容易にするために、本物の送信者が、URL600のこの部分にログイン名およびパスワードなど認証情報を組み込む可能性がある。

【0065】

正規化および明瞭化した後に、URL600は、本質的にx. y. com/a/b/cとして表すことができ、ただし、x. y. com630は装置の名前（FQDN）であり、a/b/c（たとえば、接尾語）はその装置上のファイルの位置である。x. y. com/a/b/c600がスパム送信者を識別する場合には、x. y. com/a/b610およびx. y. com/a620もまた、同じまたは関連するスパム送信者を識別する可能性が非常に高い。したがって、URL600の末端部分またはパスウェイは1度に1つの部分が分解され、たとえば、機械学習システムまたはリスト用の追加の特徴を得ることができる。これにより、すべてが実際にパターンに気付かれないような方法でスパム送信者につながる多数の様々な配置をスパム送信者が作成するのはより困難になる。

【0066】

接尾語が分解されたとき、図4において上記で説明したように、FQDN630をさらに構文解析し、追加の特徴を得ることができる。さらに、上記で図3において示されているように、FQDN630もまた、IPアドレスに変換することができる。したがって、IPアドレスに関連する様々な特徴もまた、特徴として使用することができる。

【0067】

いくつかのURLは、nnn.nnn.nnn.nnn/a/b/cなど、FQDNでなくIPアドレス（たとえば、ドット付き10進フォーマット）で記述される。接尾語は「c」で始まる連続する順番で除去することができ、各段階で、得られる（部分的な）URLを特徴として使用することができる（たとえば、nnn.nnn.nnn.nnn/a/b、nnn.nnn.nnn.nnn/a、nnn.nnn.nnn.nnnは、すべてドット付き10進フォーマットのURLから抽出することが可能な特徴である）。引き続き、（たとえば、接頭語と接尾語がない）IPアドレスを特徴として使用することができる。次いで、そのIPアドレスをそのネットブロックにマップすることができる。ネットブロックが確認できない場合には、IPアドレスの最初の1、2～最初の31ビットまでの各々を別の特徴として使用し、複数回推測することができる（図3参照）。

【0068】

ドット付き10進フォーマットに加えて、IPアドレスは、dword（ダブルワード）フォーマット（たとえば、それぞれ10進法で2つの16ビットからなる2進ワード）、8進フォーマット（たとえば、8進法）、16進フォーマット（たとえば、16進法）で表すことができる。実際には、スパム送信者は、たとえば、ドメイン名部分を%nn表記（nnは16進数字の対）を使用して符号化することにより、IPアドレス、URL、mailto:リンク、および/またはFQDNを不明瞭化する可能性がある。

【0069】

いくつかのURLは、ユーザを混乱させ、または騙すために使用することができるリダイレクタを含む可能性がある。リダイレクタは、URLのIPアドレス内で「?」に続くパラメータまたはパラメータのセットであり、別のウェブページに向き直すようにブラウザに指令する。たとえば、URLは「www.intendedpage.com?www.actualpage.com.」として現れる可能性があり、ブラウザは実際に「www.actualpage.com」を指しており、予想されている「www.intendedpage.com」ではなくそのページをロードする。したがって、URL内に含まれるパラメータもまた、特徴として抽出するために考慮することができる。

【0070】

次に、本発明による様々な方法について、一連の動作を介して説明する。いくつかの動作は、本発明に従って様々な順序で、および/または、本明細書に示され説明されるものからの他の動作と同時に進行することができるため、本発明は、動作の順序によって制限されないことを理解されたい。たとえば、方法は、これに代えて、状態図など一連の相互に関係のある状態または事象として表すことができることを、当業者なら理解できるであろう。さらに、本発明による方法を実施するのに図に示される動作のすべてが必要とされるわけではない。

【0071】

図7は、本発明の態様によるフィルタをトレーニングすることを容易にする例示的なプロセス700のフローチャートを示す。プロセス700は、プロセス710でメッセージ（たとえば、少なくとも1つのメッセージ）を受信することで開始することができる。メッセージは、たとえば、サーバによって受信されると、サーバ部の既存のフィルタ（たとえば、スパムフィルタ）は、少なくとも部分的にはフィルタによって既に学習された1組の基準に基づいて、そのメッセージをスパムの可能性があるものと、スパムの可能性がないものとともに分類することができる。プロセス720において、メッセージを構文解析し、そこから1つまたは複数の特徴を抽出する。特徴の抽出については、（図11において下記の）プロセス725でさらに詳しく説明する。特徴の例には、received f

romフィールド、reply-toフィールド、ccフィールド、mailto:フィールド、MAIL FROM SMTPコマンド、HELOフィールド、テキスト内に、または画像として埋め込まれたURLアドレス、および/または非無料電話番号（たとえば、地理的領域をマップするためのエリアコード）、ならびにメッセージ本文内のテキスト内に位置する情報（たとえば、送信者のIPアドレス）が含まれる。

【0072】

抽出された（および/または正規化された）特徴ならびにメッセージの分類（たとえば、スパムまたは非スパム）を、プロセス730でトレーニング用データセットに追加することができる。プロセス740で、上記（たとえば、プロセス710、プロセス720、プロセス730）は、実質的にすべての他の着信メッセージについて、それらがそれに応じて処理されるまで繰り返すことができる。プロセス750で、有用であると思われる特徴、または最も有用な特徴をトレーニング用セットから選択することができる。そのような選択された特徴を使用し、プロセス760で、たとえば機械学習アルゴリズムにより、機械学習フィルタなどフィルタをトレーニングすることができる。

【0073】

機械学習フィルタは、トレーニングされた後に、図8の例示的な方法800によって説明されるように、スパム検出を容易にするために使用することができる。方法800は、810でメッセージを受信することで開始される。プロセス820で、図11に関連して以下で説明するように、1つまたは複数の特徴がメッセージから抽出される。プロセス830で、抽出された特徴が、たとえば、機械学習システムによってトレーニングされたフィルタを通過する。引き続いて、「スパム」「非スパム」などの判定、またはメッセージがスパムである確率が、機械学習システムから得られる。メッセージの内容に関して判定が得られた後に、適切な措置をとることができる。措置の型には、それだけには限らないが、メッセージを削除すること、メッセージを特別なフォルダに移動すること、メッセージを隔離すること、受信者がメッセージにアクセスできるようにすることが含まれる。

【0074】

またはこれに代えて、メッセージから抽出された特徴と共に、リストをベースとする活動を実施することができる。図9は、少なくとも部分的には、抽出された特徴と、スパムまたは非スパム（あるいは、スパムである可能性が高い、または可能性が低い）と分類された受信メッセージ内でのその発生に基づいてリストを構築し、かつリストに配置するための例示的なプロセス900のフローチャートを示す。プロセス900は、プロセス910で、メッセージを受信することによって開始される。引き続いて、プロセス920で、たとえばメッセージ送信者のIPアドレスなど、いくつかの注目される特徴が抽出される。メッセージが受信された後のある時点で、メッセージを、たとえば、既存のフィルタによってスパムまたは非スパムと分類することができる。プロセス930で、メッセージの分類（たとえば、スパムまたは非スパム）に従って、特徴を増分によって計数することができる。プロセス940で、実質的にすべてのメッセージが（たとえば、プロセス910、プロセス920、プロセス930で）処理されるまで繰り返すことができる。その後、プロセス950で、特徴のリストを作成することができる。たとえば、90%問題なし（たとえば、その時点の90%非スパム、または着信メッセージの90%で非スパム）である送信者IPアドレスについて、あるリストを作成することができる。同様に、90%悪質（スパム）である送信者IPアドレスについて別のリストを作成することができる。同様の方法で、他の特徴について他のリストを作成することができる。

【0075】

これらのリストは動的に更新することができることを理解されたい。すなわち、これらのリストは、新しいメッセージの追加のグループが処理されたとき更新することができる。したがって、ある送信者のIPアドレスが、当初、問題のないリスト上で見つかり、次いで、その後ある時点で、悪質リスト上で見つかる可能性がある。というのは、一部のスパム送信者が（たとえば、フィルタならびに受信者の「信頼」を得るために）最初に問題のないメールを送信し、次いで、実質的にスパムだけを送信し始めることが普通だからで

ある。

【0076】

これらのリストは、様々な方法で使うことができる。たとえば、フィルタをトレーニングするために機械学習システムによって使用されるトレーニング用セットを生成するために使うことができる。そのようなものは、図10で次に説明される例示的なプロセス1000によって示されている。図10によれば、プロセス1000は、プロセス1010でメッセージを受信することによって始めることができる。メッセージを、たとえば、スパムまたは非スパムと分類することができる。プロセス1020で、それだけには限らないが、送信者のIPアドレスを含む特徴をメッセージから抽出することができる。プロセス1030で、抽出された特徴およびメッセージの分類がトレーニング用セットに追加され、このトレーニング用セットは、後に機械学習システムをトレーニングするために使用される。

【0077】

引き続きプロセス1040で、その送信者IPアドレスが存在する特定のリストに対応する特別な特徴がトレーニング用セット内に含まれる。たとえば、その送信者IPアドレスが「90%問題なし」リスト上にあるならば、トレーニング用セットに追加された特徴は「90%問題なしリスト」となるはずである。プロセス1050で、先行するステップ（たとえば、プロセス1010、プロセス1020、プロセス1030、プロセス1040）は、実質的にすべての着信メッセージが処理されるまで繰り返すことができる。いくつかの特徴は、フィルタをトレーニングする目的にとって他の特徴より有用である可能性があるため、プロセス1060で、最も有用な特徴または特徴群が、部分的にはユーザー嗜好に基づいて選択され、機械学習アルゴリズムを使用して、スパムフィルタなどフィルタをトレーニングするために使用される。

【0078】

さらに、たとえば、テストメッセージ、新しいメッセージ、および／または疑わしいメッセージと比較するために、IPアドレスの動的リストを構築することができる。しかし、この例では、IPアドレス自体は特徴ではない。むしろIPアドレスの品質が特徴である。またはこれに代えて、またはこれに加えて、リストを他の方法で使うことができる。実際には、たとえば、疑わしいIPアドレスのリストは、悪質であるとして送信者にフラグし、それに応じて、それらのメッセージを疑いながら処理するために使うことができる。

【0079】

次に、図11に参照すると、図7～10においてそれぞれ上記で説明されるプロセス700、プロセス800、プロセス900、プロセス1000に関連して、メッセージから特徴を抽出する例示的な方法1100のフローチャートが示されている。方法1100は、発信元IPアドレスまたはその一部分を抽出し、プロセス1110で正規化して開始することができる。また、プロセス1110で、その発信元IPアドレスから追加の特徴を抽出するため、そのIPアドレスをビット毎（たとえば、図3で説明しているように、最初の1ビット、最初の2ビット～最初の31ビットまで）の処理にかけることができる。さらに、主張されている送信者のホスト名もまた、プロセス1110で抽出することができる。次に、正規化された発信元IPアドレスおよび送信者ホスト名の特徴群を、機械学習システムまたは関連トレーニング用システムの特徴群として使うことができる。

【0080】

オプションであるが、プロセス1120にて、「From:」行の内容を抽出および／または正規化し、後に特徴として使うことができる。プロセス1130で、同様に「MAIL FROM SMTP」コマンドの内容を、特徴として使うために抽出および／または正規化することができる。

【0081】

次いで、方法1100の処理を行い、メッセージ内に含まれる可能性のある他の可能な特徴を捜すことができる。たとえば、オプションで、（必要な場合）プロセス1140に

てreply-toフィールド内の内容を抽出および正規化することができる。プロセス1150で、ccフィールドの内容を、少なくとも1つの特徴として使用するためにオプションで抽出および／または正規化することができる。プロセス1160で、非無料電話番号を、オプションでメッセージの本文から抽出し、特徴として割り当てることもできる。非無料電話番号は、その電話番号のエリアコードおよび／または最初の3桁を使用してスパム送信者の位置をマッピングすることができるため、スパム送信者を識別するのに有用となる。複数の非無料電話番号がメッセージ内に存在する場合、プロセス1160で、各番号を別の特徴として抽出および使用することができる。

【0082】

同様に、オプションで、1つまたは複数のURLおよび／またはmailto:リンク、あるいはその一部分を、それぞれプロセス1170およびプロセス1180にて抽出および／または正規化することができる。具体的には、URLをパスウェイ分解（たとえば、URLのファイル名部分）にかけることができ、URLのFQDN部分の末端に付けられている1つまたは複数の接尾語を分解することができる。これにより、パスウェイ内の接尾語の数に応じて、1つまたは複数の部分URLが得られる。各部分URLは、本発明による別の特徴として使用することができる。

【0083】

方法1100は、引き続きメッセージの本体を走査し、本物のメッセージ内よりスパムメッセージ内で見つかる可能性の高い、またその逆である他の電子メールアドレス、ならびに（たとえば、先に選択されている、または決定されている）キーワードおよび／または句を捜すことができる。各単語または句は、機械学習システム、またはリストの要素、あるいはその両方のための特徴として抽出および使用することができる。

【0084】

先に説明しているように、インターネットを介して送信されるメッセージは、必要なサーバがわずか2つでもサーバからサーバに送信される可能性がある。メッセージを処理するサーバの数は、ファイアウォールと関連ネットワークアーキテクチャが存在する結果増大する。メッセージがサーバからサーバに渡されると、各サーバは、そのIPアドレスをreceived-fromフィールドに付加する。また、各サーバは、先に付加されたどの発信元アドレスも修正する機能を有している。残念ながら、スパム送信者はこの機能を利用することができ、その位置および／または正体を偽装するため、およびメッセージの送信源について受信者を欺くため、偽物のアドレスをreceived-fromフィールドに入力することができる。

【0085】

図12は、着信メッセージのreceived-from行内で本物と偽物（たとえば、スパム送信者）の付加サーバIPアドレスを区別するための例示的なプロセス1200のフローチャートを示す。付加されたreceived-fromアドレスは、追加された順序で調べることができる（たとえば、最初のものが最も最近追加されたものである）。したがって、プロセス1210で、ユーザは、一連の送信側サーバIPアドレスを遡って追跡し、最後の信頼されるサーバIPアドレスを決定することができる。プロセス1220で、最後の信頼されるサーバIPアドレス（システムのすぐ外側のもの）を、機械学習システムによって使用される特徴として抽出することができる。最後の信頼されるものの以後の他のIPアドレスは、どれも疑わしい、または信頼できないものと見なすことができ、無視することができるが、（概ね）問題のないIPアドレスのリストおよび（概ね）悪質なIPアドレスのリストに比較することも可能である。

【0086】

プロセス1230で、主張されている送信者のFQDNもまた、その送信者が本物であるか、それともスパム送信者であるか判定することを容易にするために抽出することができる。より具体的には、主張されているFQDNをドメイン分解によって分析し、複数の部分FQDNに細分化することができる。たとえば、主張されているFQDNがa. b. c. x. comであると想像する。この主張されているFQDNは、b. c. x. com

→c. x. com→x. com→comを生成するような形で分解されるはずである。したがって、各部分FQDNセグメントならびに完全FQDNを別の特徴として使用し、偽物の送信者と本物の送信者を決定する際に助けとすることができる。

【0087】

本発明はまた、ペアレנטル制御システムを使用することができる。ペアレントル制御システムは、少なくとも部分的には、メッセージの一部の内容に基づいて、見るのに不適切なものとしてメッセージを分類し、不適切分類の理由を提供することができる。たとえば、(テキストまたは画像をベースとする)クリック可能なリンクとして、またはメッセージの本文内のテキストとして、URLがメッセージ内に埋め込まれている可能性がある。ペアレントル制御システムは、埋め込まれたURLと、その記憶された問題なし、および/または悪質URLリストの1つまたは複数とを比較し、あるいはペアレントル制御分類のための他の技法を使用して、メッセージの適切な分類を決定することができる。次いで、この分類は、機械学習システム内で、または特徴リスト上で、あるいはその両方で追加の特徴として使用することができる。

【0088】

図13では、ペアレントル制御システムの少なくとも1つの態様を本発明に組み込むための例示的なプロセス1300のフローチャートが示されている。プロセス1310で1組のメッセージを受信した後に、プロセス1320で、URL、mailto:リンクもしくはmailto:リンクに類似の他のテキスト、またはURLの一部分を求めてそのメッセージを走査することができる。プロセス1330で、メッセージが上記のどれかを含んでいるようには考えられない場合、プロセス1300は、プロセス1310に戻る。しかし、メッセージがそのように考えられる場合、プロセス1340で、検出された特徴の少なくとも一部分を少なくとも1つのペアレントル制御システムに渡すことができる。

【0089】

プロセス1350で、ペアレントル制御システムは、URL、mailto:リンク、URLサービス名、URLパス、FQDN(たとえば、URL、電子メールアドレスなどのFQDN部分など)の1つまたは複数のデータベースを調べることにより、mailto:リンク、URL、またはそれらの一部分を分類することができる。たとえば、メッセージを、ポルノ、「借金返済」、賭博、および他の同様な題材の少なくとも1つを含むものとして分類することができる。そのような分類は、プロセス1360で追加の特徴として抽出することができる。スパムメッセージの大部分の内容はそのような題材を含むため、ペアレントル制御システムを組み込むことは、機械学習システムがトレーニングを行い、改良されたフィルタを構築するためにさらに使用することができる特徴を得る上で有用となる。他の分類も考えられるが、憎悪発言、性的題材、銃暴力、および麻薬関連題材も含まれ、そのような分類もまた、特徴として使用することができるものの、これには限られない。スパムメッセージは、これらの種類の題材に関連する内容を含むこともあり、含まないこともあるが、ユーザは、依然としてこれらの種類のメッセージを遮断したいと望む可能性がある。

【0090】

実際には、様々な分類により、様々なスパム度を示すことができる。たとえば、憎悪発言として分類されたメッセージは、(たとえば、スパムでない可能性が最も高いため)実質的にスパム度を示さない可能性がある。逆に、性的内容/題材として分類されたメッセージは、比較的高いスパム度を反映している可能性がある(たとえば、メッセージがスパムである90%以下の確実性)。機械学習システムは、スパム度を反映するフィルタを構築することができる。したがって、ユーザ嗜好を満たすようにフィルタをカスタマイズおよび個別化することができる。

【0091】

すでに説明したように、無数の特徴をメッセージから抽出し、機械学習システムによるトレーニング用データとして、または、問題のない特徴および悪質な特徴を識別するリスト上の要素として使用することができる。特徴自体に加えて、特徴の質は、スパムを検出

および防止する上で有用となる。たとえば、ある特徴が送信者の電子メールアドレスであると考え、その電子メールアドレスをある特徴として使用することができ、新しい着信メッセージ内に現れるその電子メールアドレスの頻度または回数を別の特徴として使用することができるであろう。

【0092】

図14は、この型の（たとえば、抽出された特徴の共通性または希少性に関する）特徴を抽出するための例示的なプロセス1400のフローチャートを示す。スパム送信者は、その位置を急いで変更しようと頻繁に試み、その結果、たとえば以前見られなかったアドレスから、または以前は知られていなかった装置を指すURLを有するメールを送信する可能性がほとんどのユーザより高い。したがって、抽出された各特徴の型（たとえば、発信元IPアドレス、URL、電子メールアドレス、ドメイン名など）について、各型の特徴のリストが保持されているものと仮定して、特定の特徴の発生率、頻度、または回数を追跡することができる。

【0093】

プロセス1400は、プロセス1410で、着信メッセージからの1つまたは複数の特徴の抽出、および／または特徴の正規化から開始する。次いで、プロセス1420でその特徴を、先に抽出された、または複数の以前のメッセージ内で観察された特徴の1つまたは複数のリストに比較することができる。次いで、プロセス1430は、現在の特徴が共通であるかどうか判定することができる。特徴の共通性は、最近の、および／または以前の着信メッセージ内にその特徴が現れる頻度を計算することによって決定することができる。プロセス1430で、そのメッセージが共通でない、またはそれほど共通でない（たとえば、共通性閾値を満たしていない）場合には、プロセス1440で、その希少性を追加の特徴として使用することができる。そうでない場合には、プロセス1460で、その特徴の共通性もまた特徴として使用することができる。

【0094】

上述した本発明によれば、以下の擬似コードを使用し、本発明の少なくとも1つの態様を実施することができる。変数名は、すべて大文字で示されている。追加の注意として、`add-machine-features`と`add-ip-features`という2つの関数が擬似コードの末尾で定義されている。「`PREFIX-machine-MACHINE`」のような表記を使用し、`PREFIX`変数内にあるものが「`machine`」という単語に連結され、さらに「`machine`」という単語が`MACHINE`変数内にあるものに連結されて構成されたストリングを示す。最後に、関数`add-to-feature-list`は、現在のメッセージに関連付けられた特徴のリストに特徴を書き込む。

【0095】

例示的な擬似コードは、次の通りである。

【0096】

【表5】

```
# for a given message, extract all the features
```

```
IPADDRESS := the last external IP address in the received-  
from list;
```

【0097】

【表6】

```

add-ipfeatures(received, IPADDRESS);
SENDERS-ALLEGED-FQDN := FQDN in the last external IP
address in the received-from list;
add-machine-features(sendersfqdn, SENDERS-ALLEGED-FQDN);

for each email address type TYPE in (from, CC, to, reply-
to, embedded-mailto-link, embedded-address, and SMTP MAIL
FROM)
{
    for each address ADDRESS of type TYPE in the message {
        deobfuscate ADDRESS if necessary;
        add-to-feature-list TYPE-ADDRESS;
        if ADDRESS is of the form NAME@MACHINE then
        {
            add-machine-features(TYPE, MACHINE);
        }
        else
        { # ADDRESS is of form NAME@IPADDRESS
            add-ip-features(TYPE, IPADDRESS);
        }
    }
}

for each url type TYPE in (clickable-links, text-based-
links, embedded-image-links)
{
    for each URL in the message of type TYPE
    {
        deobfuscate URL;
        add-to-feature-list TYPE-URL;
        set PARENTALCLASS := parental control system class
of URL;
        add-to-feature-list TYPE-class-PARENTCLASS;
        while URL has a location suffix
        {
            remove location suffix from URL, i.e. x.y/a/b/c
-> x.y/a/b; x.y/a/b -> x.y/a; x.y/a;
        }
        # All suffixes have been removed; URL is now either
machine name or IP address
        if URL is machine name
        {
            add-machine-features(TYPE, URL);
        }
        else
        {
            add-ip-features(TYPE, URL);
        }
    }
}

```

【表7】

```

    }
  }
}

function add-machine-features(PREFIX, MACHINE)
{
  add-ip-features(PREFIX-ip, nslookup(MACHINE));
  while MACHINE not equal ""
  {
    add-to-feature-list PREFIX-machine-MACHINE;
    remove beginning from MACHINE # (i.e. a.x.com ->
x.com, or x.com -> com);
  }
}

function add-ip-features(PREFIX, IPADDRESS)
{
  add-to-feature-list PREFIX-ipaddress-IPADDRESS;
  find netblock NETBLOCK of IPADDRESS;
  add-to-feature-list PREFIX-netblock-NETBLOCK;
  for N = 1 to 31 {
    MASKED = first N bits of IPADDRESS;
    add-to-feature-list PREFIX-masked-N-MASKED;
  }
}

```

【0099】

本発明の様々な態様についてさらに情報を提供するため、図15および以下の考察では、本発明の様々な態様を実施することができる好適な動作環境1510を簡単に、一般的に説明するものとする。本発明について、1つまたは複数のコンピュータまたは他のデバイスによって実行されるプログラムモジュールなどのコンピュータ実行可能命令の一般的な文脈で説明されるが、本発明はまた、他のプログラムモジュールとの組合せで、および／またはハードウェアとソフトウェアの組合せとして実施することができることを、当業者であれば理解することができるであろう。

【0100】

ただし、一般に、プログラムモジュールは、特定のタスクを実行する、あるいは特定の抽象データ型を実施するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。動作環境1510は、好適な動作環境の一例にすぎず、本発明の使用または機能の範囲についてどんな制限も暗示しないものとする。本発明と共に使用するのに適している可能性のある他の周知のコンピュータシステム、環境、および／または構成には、それだけには限らないが、パーソナルコンピュータ、ハンドヘルドデバイスまたはラップトップデバイス、マルチプロセッサシステム、マイクロプロセッサをベースとするシステム、プログラム可能な家電、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、上記のシステムまたはデバイスを含む分散コンピューティング環境などが含まれる。

【0101】

図15を参照すると、本発明の様々な態様を実施するための例示的な環境1510は、コンピュータ1512を含んでいる。コンピュータ1512は、処理装置1514、シス

テムメモリ1516、システムバス1518を含む。システムバス1518は、それだけには限らないが、システムメモリ1516を含むシステム構成要素を処理装置1514に結合する。処理装置1514は、様々な使用可能なプロセッサのいずれかとしてすることができる。デュアルマイクロプロセッサおよび他のマルチプロセッサアーキテクチャもまた、処理装置1514として使用することができる。

【0102】

システムバス1518は、メモリバスもしくはメモリコントローラ、周辺機器バスもしくは外部バス、および／または任意の様々な使用可能なバスアーキテクチャを使用するローカルバスを含めて、いくつかの型のバス構造のいずれかとしてすることができる。バスアーキテクチャには、それだけには限らないが、11ビットバス、ISA (Industrial Standard Architecture) バス、MCA (Micro-Channel Architecture) バス、EISA (Extended ISA) バス、IDE (Intelligent Drive Electronics)、VESAローカルバス (VLB)、PCI (Peripheral Component Interconnect)、ユニバーサルシリアルバス (USB)、AGP (Advanced Graphics Port)、PCMCIA (Personal Computer Memory Card International Association) バス、SCSI (Small Computer Systems Interface) が含まれる。

【0103】

システムメモリ1516には、揮発性メモリ1520および不揮発性メモリ1522が含まれる。起動中などにコンピュータ1512内の要素間で情報を転送する基本ルーチンを含む基本入出力システム (BIOS) は、不揮発性メモリ1522内に記憶されている。限定ではなく例を挙げると、不揮発性メモリ1522には、読み出し専用メモリ (ROM)、プログラム可能なROM (PROM)、電気的プログラム可能なROM (EPROM)、電気的消去可能なROM (EEPROM)、またはフラッシュメモリが含まれる。揮発性メモリ1520には、外部キャッシュメモリとして動作するランダムアクセスメモリ (RAM) が含まれる。限定ではなく例を挙げると、RAMは、スタティックRAM (SRAM)、ダイナミックRAM (DRAM)、シンクロナスDRAM (SDRAM)、ダブルデータレートSDRAM (DDR SDRAM)、ESDRAM (enhanced SDRAM)、SLDRAM (Synchlink DRAM)、ダイレクトラムバスRAM (DRRAM) など、多数の形態で使用可能である。

【0104】

コンピュータ1512はまた、取外し可能／固定、揮発性／不揮発性コンピュータ記憶媒体を含む。図15は、たとえば、ディスク記憶装置1524を示す。ディスク記憶装置1524には、それだけには限らないが、磁気ディスクドライブ、フロッピー（登録商標）ディスクドライブ、テープドライブ、Jazドライブ、Zipドライブ、LS-100ドライブ、フラッシュメモリカード、メモリースティックのようなデバイスが含まれる。さらに、ディスク記憶装置1524には、それだけには限らないが、コンパクトディスクROMデバイス (CD-ROM)、記録可能なCDドライブ (CD-Rドライブ)、再書き込み可能なCDドライブ (CD-RWドライブ)、またはデジタル多用途ディスクROMドライブ (DVD-ROM) など光ディスクドライブを含めて、記憶媒体が別個に、または他の記憶媒体との組合せで含まれる可能性がある。ディスク記憶装置1524のシステムバス1518に対する接続を容易にするために、インターフェース1526など取外し可能または固定インターフェースが一般に使用される。

【0105】

図15は、ユーザと、好適な動作環境1510に説明される基本的なコンピュータ資源との間を媒介するものとして動作するソフトウェアについて述べていることを理解されたい。そのようなソフトウェアには、オペレーティングシステム1528が含まれる。オペレーティングシステム1528は、ディスク記憶装置1524に記憶することができ、コ

ンピュータシステム1512の資源を制御し、割り振るように動作する。システムアプリケーション1530は、システムメモリ1516内またはディスク記憶装置1524に記憶されたプログラムモジュール1532およびプログラムデータ1534を介して、オペレーティングシステム1528による資源の管理を利用する。本発明が様々なオペレーティングシステムまたはオペレーティングシステムの組合せと共に実施することができることを理解されたい。

【0106】

ユーザは、入力デバイス1536を介してコンピュータ1512にコマンドまたは情報を入力する。入力デバイス1536には、それだけには限らないが、マウスなどポインティングデバイス、トラックボール、スタイラス、タッチパッド、キーボード、マイクロフォン、ジョイスティック、ゲームパッド、衛星パラボラアンテナ、スキャナ、TV同調器カード、デジタルカメラ、デジタルビデオカメラ、ウェブカメラなどが含まれる。これら、および他の入力デバイスは、インターフェースポート1538を介して、システムバス1518を通じて処理装置1514に接続する。インターフェースポート1538には、たとえば、シリアルポート、パラレルポート、ゲームポート、ユニバーサルシリアルバス(USB)が含まれる。出力デバイス1540は、入力デバイス1536と同じ型のポートのいくつかを使用する。したがって、たとえばUSBポートは、コンピュータ1512に入力を提供するために、また、コンピュータ1512から出力デバイス1540に情報を出力するために使用することができる。出力アダプタ1542は、出力デバイス1540の中でも、特別なアダプタを必要とするモニタ、スピーカ、プリンタのようないくつかの出力デバイス1540があることを示すために提供されている。限定ではなく例を挙げると、出力アダプタ1542には、出力デバイス1540とシステムバス1518の間で出力手段を提供するビデオカードおよびサウンドカードが含まれる。他のデバイスおよび/またはデバイスのシステムは、リモートコンピュータ1544など、入力機能と出力機能を共に提供することに留意されたい。

【0107】

コンピュータ1512は、リモートコンピュータ1544など、1つまたは複数のリモートコンピュータに対する論理接続を使用してネットワーク環境内で動作することができる。リモートコンピュータ1544は、パーソナルコンピュータ、サーバ、ルータ、ネットワークPC、ワークステーション、マイクロプロセッサをベースとする家電、ピアデバイスまたは他の通常のネットワークノードなどとなることができ、一般に、コンピュータ1512に関して上述した要素の多数または全部を含む。簡潔にするために、メモリ記憶装置1546だけがリモートコンピュータ1544と共に示されている。リモートコンピュータ1544は、ネットワークインターフェース1548を介してコンピュータ1512に論理的に接続され、次いで、通信接続1550を介して物理的に接続される。ネットワークインターフェース1548は、ローカルエリアネットワーク(LAN)および広域ネットワーク(WAN)など、通信ネットワークを含む。LAN技術には、光ファイバ分散データインターフェース(FDDI)、銅分散データインターフェース(CDDI)、イーサネット(登録商標)/IEEE1102.3、トークンリング/IEEE1102.5などが含まれる。WAN技術には、それだけには限らないが、ポイント・トゥ・ポイント・リンク、総合デジタル通信網(ISDN)のような回路交換ネットワークとその変形形態、パケット交換ネットワーク、およびデジタル加入者回線(DSL)が含まれる。

【0108】

通信接続1550は、ネットワークインターフェース1548をバス1518に接続するために使用されるハードウェア/ソフトウェアを指す。通信接続1550は、図が見やすいようにコンピュータ1512の内側で示されているが、コンピュータ1512の外側とすることもできる。例示する目的にすぎないが、ネットワークインターフェース1548に対する接続に必要なハードウェア/ソフトウェアには、通常、電話級モデム、ケーブルモデムおよびDSLモデムを含むモデム、ISDNアダプタ、ならびにイーサネット(登録商標)カードなど、内部技術および外部技術が含まれる。

【0109】

上述したものには、本発明の諸例が含まれる。当然ながら、本発明について説明するために構成要素または方法の考えられるあらゆる組合せについて説明することは可能でなく、本発明の多数の他の組合せおよび変形が可能であることを、当業者なら理解することができる。したがって、本発明は、添付の特許請求の範囲の精神および範囲内に入るそのような変更、修正、および変形形態をすべて包含するものとする。さらに「include（含む）」という用語が詳細な説明または特許請求の範囲で使用されている限り、そのような用語は、「comprising（含む、備える）」が特許請求の範囲内で転換句として使用されたとき解釈されるように「comprising」という用語と同様に包括的であるものとする。

【図面の簡単な説明】

【0110】

【図1】本発明の態様による、スパム防止を容易にするシステムの高レベルブロック図である。

【図2】本発明の態様による、着信メッセージから1つまたは複数の特徴を抽出することによってスパム防止を容易にするシステムのブロック図である。

【図3】本発明の態様による、IPアドレスから抽出することができる複数の特徴の概略図である。

【図4】本発明の態様による、FQDNから抽出することができる複数の特徴の概略図である。

【図5】本発明の態様による、電子メールアドレスから抽出することができる複数の特徴の概略図である。

【図6】本発明の態様による、URLまたはウェブアドレスから抽出することができる複数の特徴の概略図である。

【図7】本発明の態様による、フィルタをトレーニングすることに関連する例示的な方法のフローチャートである。

【図8】本発明の態様による、トレーニングされたフィルタを使用することに関連する例示的な方法のフローチャートである。

【図9】本発明の態様による、リストを作成することに関連する例示的な方法のフローチャートである。

【図10】本発明の態様による、フィルタをトレーニングするためにリストを使用することに関連する例示的な方法のフローチャートである。

【図11】本発明の態様による、少なくとも図7および図8の方法の中で参照された工程のフローチャートである。

【図12】本発明の態様による、本物と偽物の発信元(received-from) IPアドレスを区別することを容易にする工程のフローチャートである。

【図13】本発明の態様による、着信メッセージから特徴を生成および/または抽出する際にペアレンタル制御システムを組み込む方法のフローチャートである。

【図14】本発明の態様による、機械学習システム内で使用される特徴セットの作成を容易にする方法のフローチャートである。

【図15】本発明の様々な態様を実施するための例示的な環境の図である。

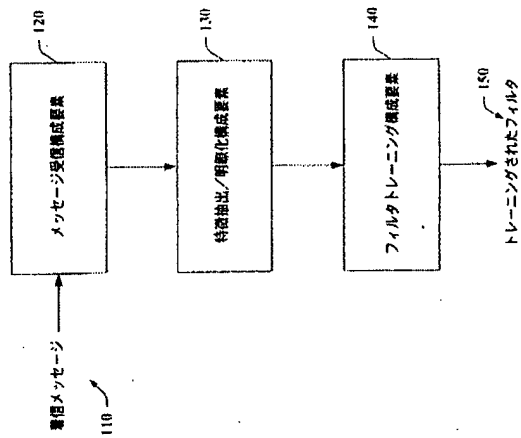
【符号の説明】

【0111】

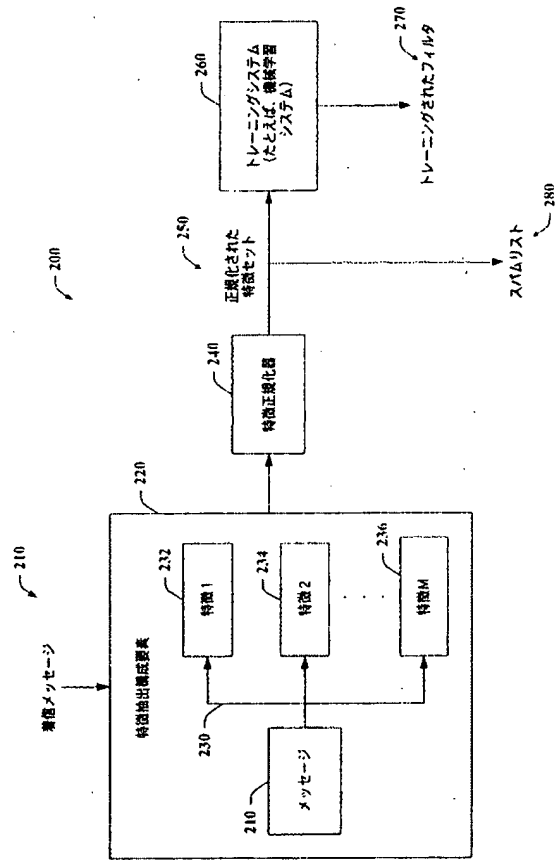
- 110 着信メッセージ
- 120 メッセージ受信構成要素
- 130 特徴抽出/不明瞭構成要素
- 140 フィルタトレーニング構成要素
- 150 トレーニングされたフィルタ
- 210 着信メッセージ、メッセージ
- 220 特徴抽出構成要素

- 230 特徴
- 232 特徴1
- 234 特徴2
- 236 特徴M
- 240 特徴正規化器
- 250 正規化された特徴セット
- 260 トレーニングシステム
- 270 トレーニングされたフィルタ
- 280 スパムリスト
- 300 IPアドレス
- 302 ブロックID
- 314 IPアドレスの階層
- 316 特徴の希少性
- 400 FQDN
- 402 ホスト名
- 404 ドメイン名
- 406 内容
- 414 特徴の型
- 500 電子メールアドレス
- 502 ユーザ名
- 504 FQDN
- 600 URL
- 1510 環境
- 1512 コンピュータ
- 1514 処理装置
- 1516 システムメモリ
- 1518 システムバス
- 1520 揮発性メモリ
- 1522 不揮発性メモリ
- 1524 ディスク記憶装置
- 1526 インターフェース
- 1528 オペレーティングシステム
- 1530 アプリケーション
- 1532 モジュール
- 1534 データ
- 1536 入力デバイス
- 1538 インターフェースポート
- 1540 出力デバイス
- 1542 出力アダプタ
- 1544 リモートコンピュータ
- 1546 メモリ記憶装置
- 1548 ネットワークインターフェース
- 1550 通信接続

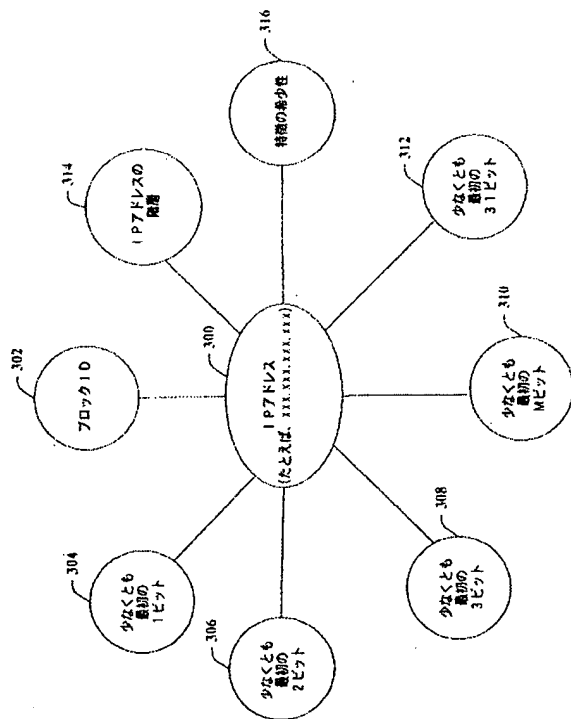
【図1】



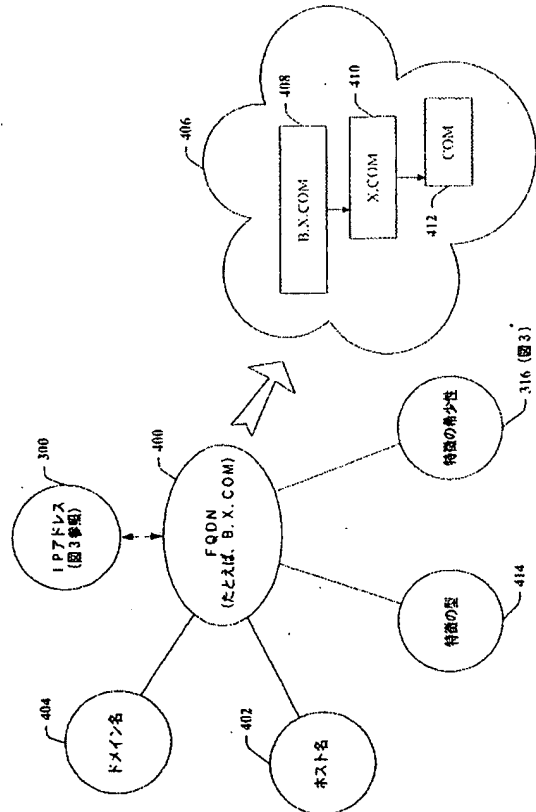
【図2】



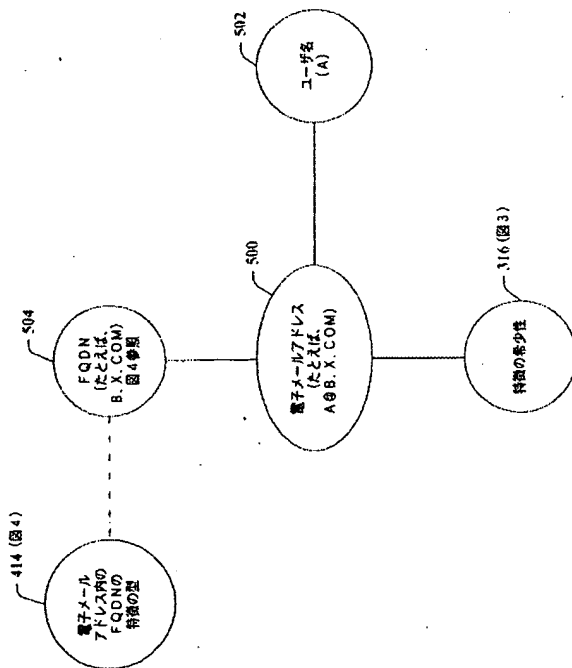
【図3】



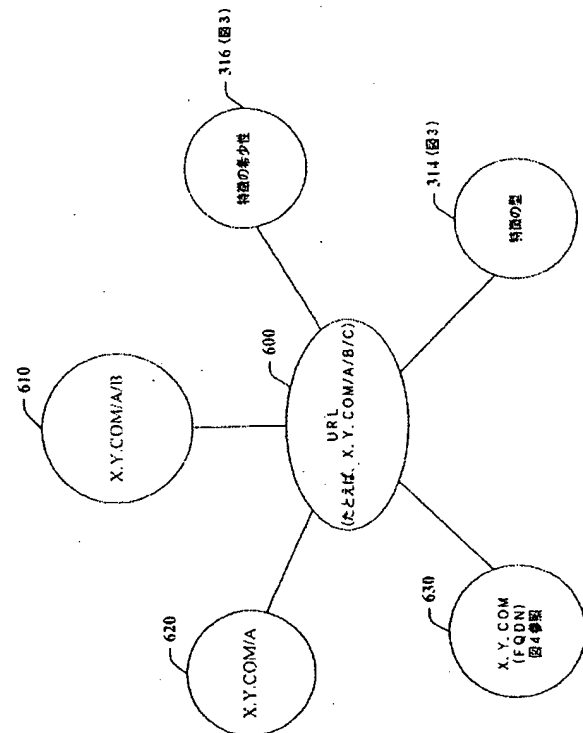
【図4】



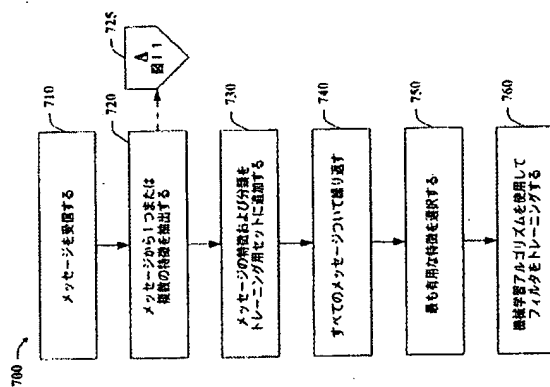
【図5】



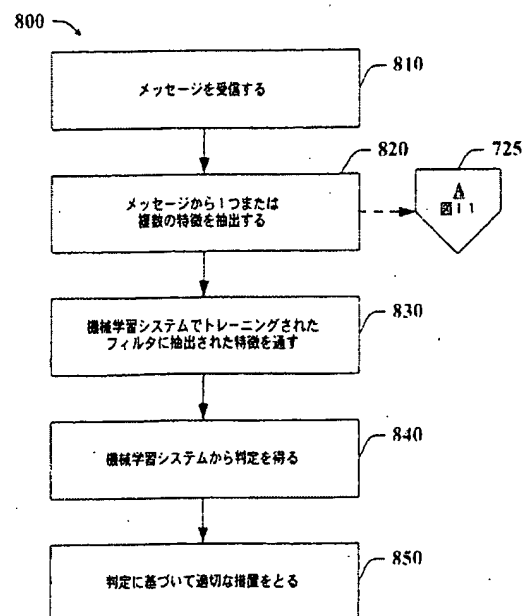
【図6】



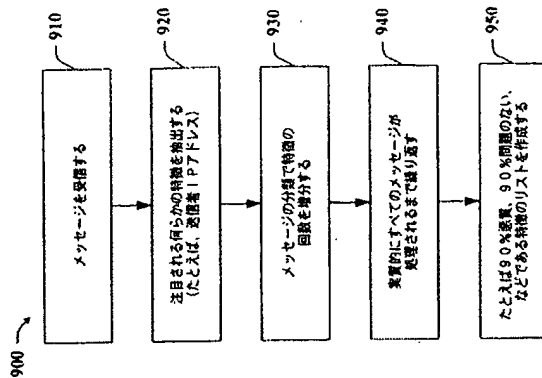
【図7】



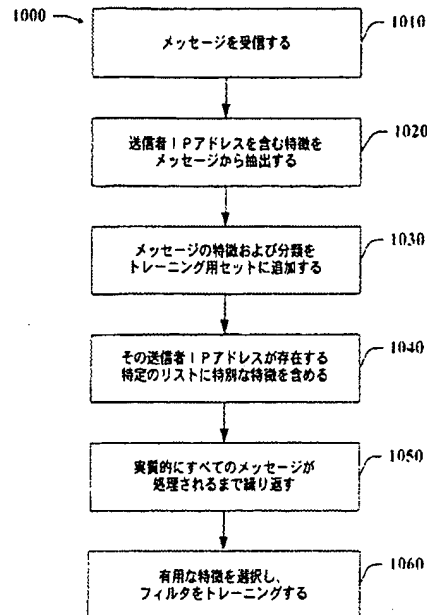
【図8】



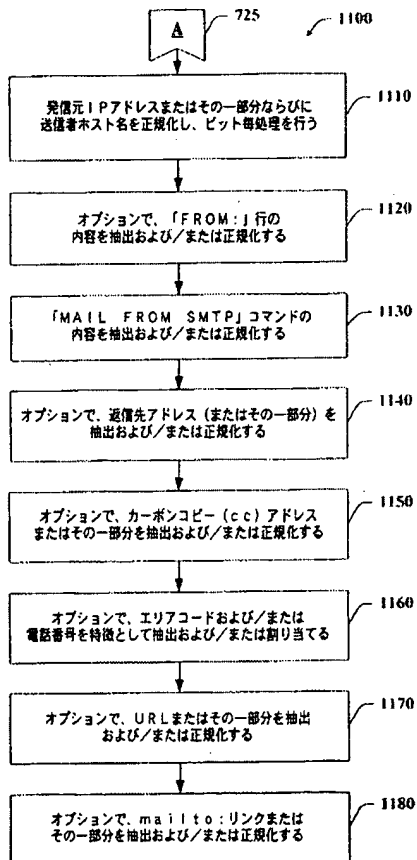
【図9】



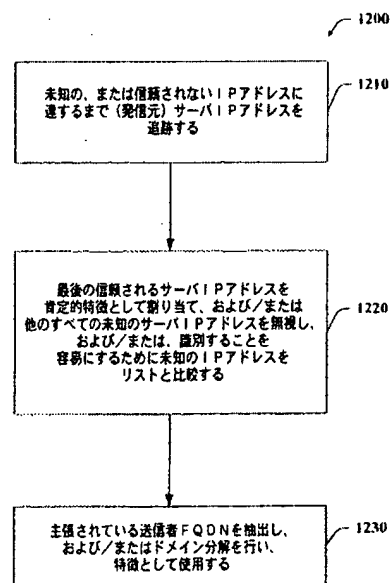
【図10】



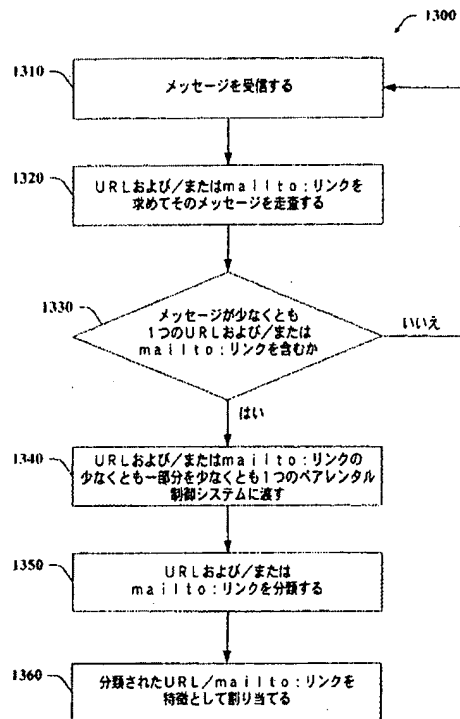
【図11】



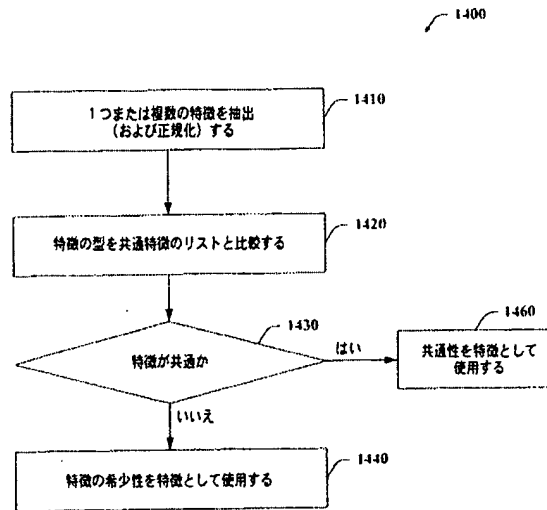
【図12】



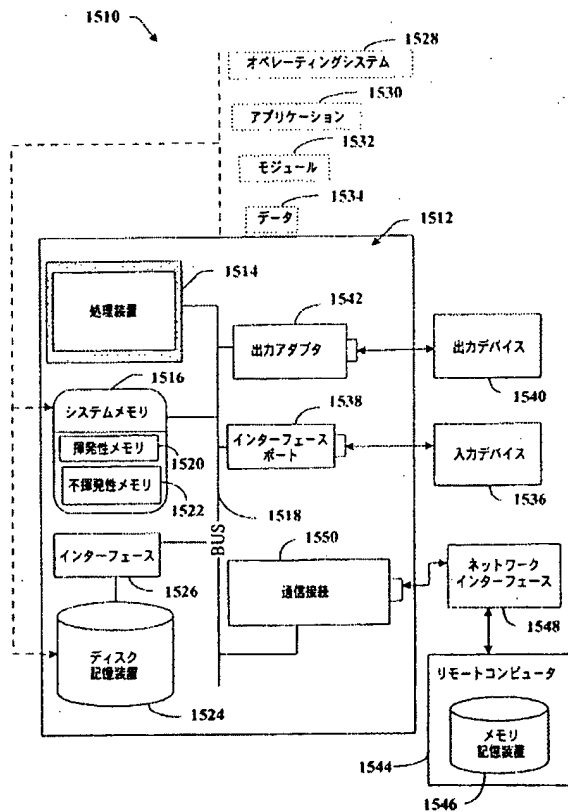
【図13】



【図14】



【図15】



- (72)発明者 ロバート エル. ラウンスウェイト
アメリカ合衆国 98024 ワシントン州 フォール シティ 287 アベニュー サウスイ
ースト 4148
- (72)発明者 ダニエル グォゼズ
アメリカ合衆国 98074 ワシントン州 サマミッシュ ノースイースト 23 コート 2
0533
- (72)発明者 ジョン ディー. メヘル
アメリカ合衆国 98103 ワシントン州 シアトル ホイットマン アベニュー ノース 3
624 ナンバー3
- (72)発明者 ネイサン ディー. ハウエル
アメリカ合衆国 98133 ワシントン州 シアトル ノース 105 ストリート 939
アパートメント エー
- (72)発明者 ミカ シー. ルパーズバーグ
アメリカ合衆国 98122 ワシントン州 シアトル イースト パイン ストリート 417
ナンバー209
- (72)発明者 ブライアン ティー. スターバック
アメリカ合衆国 98019 ワシントン州 ドボル ノースイースト 140 コート 275
17

【要約の続き】

【外国語明細書】

2004362559000001.pdf